DOCUMENT RESUME

ED 190 222 PS 011 560

AUTHOR	LaBelle, Thomas: And Others
TITLE	An Evaluation of the Head Start Bilingual Bicultural Curriculum Development Project. A Qualitative/Quantitative Data Gathering Approach (with Addendum).
INSTITUTION	Juarez and Associates, Inc., Los Angeles, Calif.
SPONS AGENCY	Administration for Children, Youth, and Families (DHEW), Washington, D.C.
PUB DATE	1 Feb 79
CONTRACT	105-77-1048
NOTE	130p.; For related documents, see PS 011 556-561, and PS 011 583.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Bilingual Education; *Curriculum Evaluation; Data Analysis; Data Collection; Ethnography; *Evaluation Methods; Multicultural Education; *Preschool Curriculum; Preschool Education; Psychometrics; Spanish Speaking
IDENTIFIERS	*Project Head Start; Qualitative Analysis; Quantitative Analysis

ABSTRACT

This 1978 report advises concerned persons and agencies of the nature and extent of plans developed by Juarez and Associates to use qualitative data in the evaluation of the Head Start Bilingual Bicultural Curriculum Development Project. Contractual responsibilities, multimethod data collection strategies, objectives and procedures are described. Additionally, evaluation procedures, data management and analysis considerations as well as the integration of psychometric and ethnographic data in statistical analysis are discussed. An addendum to the report: (1) summarizes and integrates qualitative and quantitative aspects of the evaluation design, (2) describes procedures recommended for remaining phases of the contract, and (3) provides revisions to previous reports. Related materials are appended. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190222

PS011560

Juárez and Associates, Inc.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN EVALUATION OF THE HEAD START BILINGUAL BICULTURAL CURRICULUM DEVELOPMENT PROJECT

A Qualitative/Quantitative Data Gathering Approach

Authors: Thomas LaBelle, Luis C. Moll
Ray S. Pérez, Thomas Weisner,
Marlene Zepeda

December, 1978

Contract No. HEW 105-77-1048

Project Director: Luis C. Moll, Ph.D.
Co-Project Director: Ray S. Pérez, Ph.D.

Prepared for:

Dr. Steve Martinez, Project Officer
Administration for Children, Youth,
and Families
Office of Human Development Services
Department of Health, Education, and
Welfare

1100 Glendon Avenue • Los Angeles, California 90024 • (213) 478-0826 • (213) 879-5899

Acknowledgments

A plan of this scope and complexity requires the creativity and hard work of many individuals. In particular, we would like to acknowledge the contributions of Ms. Sandra Green, who helped prepare the manuscript, and Ms. Margaret Mason, who provided editorial guidance.

TABLE OF CONTENTS

	Title	Page
I.	Introduction	I-1
	Contractual Responsibilities	I-2
	The feasibility of successfully implementing the models in more than one setting	
	The extent to which the models, once implemented, are meeting their objectives	
	The extent to which the models are greeted favorably by Head Start staff, parents, and key community members	
	Multi-Method Data Collection Strategy	I-5
	Rationale	I-5
	The need to define the treatment	
	The need to describe and define the programs' participants and setting	
	The need to describe control group activities	
	The need to observe classroom activities in various contexts	
	Objectives	I-9
	To assess the nature and extent of program implementation over time	
	To use information gathered from observations to help interpret the quantitative (test) results	
	To provide qualitative change-data over time	
	Procedures	I-11
	The Participant Researcher	I-11
	Research Strategies	I-12
	Naturalistic observations	
	Teacher interviews	
	Phone-call updates	
	Psychometric testing	
	A Preview of the Discussion	I-12
II.	Evaluation Procedures	II-1
	A Change Over Time Approach	II-1
	Study Variables	II-3
	Program Implementation	II-3
	The Setting and Material Resources	
	Individuals	
	Schedule and Organization of the Program	
	Attitudes	
	Activities	
	Program Impact	II-12
	Classroom Behaviors	
	Control Group Activities	

Overall Strategy: The Participant Researcher Role II-17
Data Collection Activities of Participant Researcher II-26
Focused systematic classroom observations
Examples
Participant researcher ratings
Examples
Ethnographic notes
Phone-call updates
Teacher interviews
Teacher ratings
Psychometric testing

III. Data Management III-1
Monitoring Data Quality III-1
Reactivity of Participant Researcher Data

IV. Data Analysis: Uses of Qualitative Notes IV-1
Domain and Indexing System
Mid-Year and Final Reviews of Field Notes
Interpretation of Quantitative Results
Exogenous Factors
Generation of New Hypotheses
Conclusion

V. Integration of Psychometric and Ethnographic Data in Statistical Analysis V-1

VI. Appendixes VI-1
Appendix A: Ethnographic Component's Participant Researcher: Task Analysis VI-1
Appendix B: Cross-Model Behaviors VI-3
Appendix C: Example of Observation Checklist VI-8
Appendix D: Job Description of Participant Researcher VI-9
Appendix E: Job Description of Coordinator of Participant Researchers VI-11

BIBLIOGRAPHY

I. Introduction

This preliminary report has been prepared to inform curriculum and model developers and the Administration for Children, Youth and Families (ACYF) staff members involved in the Evaluation of the Head Start Bilingual Bicultural Curriculum Development Project of the nature and extent of a plan being developed by Juárez and Associates for the use of qualitative data as an integral part of the evaluation design specified in the evaluation contract. This report has been requested by ACYF as a preliminary "discussion document" in the upcoming meeting (December 20, 1978) in Washington, D.C., between representatives of Juárez and Associates, the curriculum model developers, and the ACYF staff. One of the principal reasons for this meeting is to secure the input of the curriculum model developers. This is particularly important in this evaluation because the objects of study are four different curriculum models that are being implemented at a total of eight different sites. With this multiplicity of treatments, agreements as to the kind of data to be collected and the methods that can legitimately be used for that collection are central to the success of the evaluation. It is clear, for example, that changes have occurred within every model since each was originally formulated. Furthermore, additional changes are possible in the models during the remainder of this academic year. Consequently, it is important for the evaluators to work closely with the curriculum contractors to reflect these changes in the data base generated by the evaluation.

Contractual Responsibilities

This section addresses some of the issues that have shaped and focused the integrated approach that characterizes this evaluation. Whether through qualitative or quantitative means, the first characteristic is that this design requires a process of actively and consciously directed inquiry. Hence, whether through standardized instruments or observations, decisions about what questions are to be answered and what relevant behaviors are to be studied are being made before the researcher is in the field. These decisions are primarily guided by what is stipulated in the evaluation contract.

The work statement in the evaluation contract with Juárez and Associates specifies that the purpose of the evaluation is to assess the effectiveness of four early childhood bilingual bicultural curriculum models for Spanish speaking children. More specifically, the contract requires the evaluators to collect information on the following:

(1) The feasibility of successfully implementing the models in more than one setting. That is to say, the evaluators must secure information that will assist others in learning about the potential of the model for implementation elsewhere. This activity requires the evaluators to collect information regarding both the process necessary to implement each model and the procedures needed to maintain each model in a new environment, including descriptions of any special characteristics of Head Start staff, students, parents, resources, and/or community needed to assure success in its implementation.

(2) The extent to which the models, once implemented, are meeting their objectives. This means that one emphasis in this evaluation is on measuring the change in children, teachers, and parents as a result of their participation in the four curriculum models. In this regard the contract specifies an experimental pre- and post-test design. Children are to be assigned randomly to treatment and control (no-treatment) groups. Subsequently, at the beginning and conclusion of the treatment, they are tested on selected competency measures (see Juárez and Associates, July 1978). This testing may include the following domains:

Spanish language comprehension

Spanish language production

English language comprehension

English language production

Concept development

Socioemotional behaviors

A questionnaire will be administered to parents which is intended to assess background characteristics, attitudes, and knowledge with regard to education in general and bilingual education in particular, and attitudes toward educational and vocational careers for their children. Head Start teachers will complete a questionnaire designed to assess their background, their understanding of the terms "bilingual" and "bicultural" in the context of an early childhood program, their attitudes toward children, their sensitivity to special

ethnic and linguistic characteristics of children, and their ability to use these characteristics in a positive fashion in the instructional program.

(3) The extent to which the models are greeted favorably by Head Start staff, parents, and lay community members. This means collecting information regarding staff satisfaction with their participation in implementing each model; parents' attitudes on whether the particular model reflects their interests for their children and their children's needs; staff opinions as to the degree of clarity of the curriculum package and the training modules, how practical the curriculum contents, lesson plans, etc., are, and the ability of the models to take into account the specific needs of their program and the characteristics of the environment in which their program is located.

Finally, a set of pamphlets have to be developed that summarize all the information that is to be collected above. The purpose of the pamphlets is to aid in the dissemination of the four models to interested Head Start programs. Therefore, the pamphlets will include descriptions of the models, implementation information, assessment of the specific strengths of each model, and information regarding the impact of each model on parents, children, and teachers.

Multi-Method Data Collection Strategy

Rationale

In light of these evaluation goals, Juárez and Associates has repeatedly emphasized the need to expand the current experimental approach (i.e., pre- and post-testing) in order to take into account, among other factors, the actual school/classroom processes (treatment) which intervene between program input (children's characteristics, attitudes, etc.) and outcomes (dependent variables). In so doing we have advocated the use of several data-gathering methods, beyond standardized tests of a psychometric type, to include observations, ratings, event sampling, and ethnographic notes. The following evaluation-related issues make the necessity of using several data-gathering methods even more apparent.

(1) The need to define the treatment. In the present evaluation four different curriculum models are being studied. These models vary in length of existence, reasons for existing, educational philosophy and approach, training of teachers, and, of course, strategy for implementation. Although these models share several important outcome goals (e.g., to increase second language comprehension), it is clear that the process by which to achieve these goals differentiates each program.

In addition to this non-uniformity of treatments, these four curriculum models will be implemented at two sites each (for a total of eight evaluation sites) that differ in their location, in their demographic makeup, in their reasons for wanting the curriculum projects, and in their relationship with the model's staff. Since several classrooms may be involved at each site, we

can expect variability in treatment, both within and across replication sites. Cummins (1977) pointed out that evaluations which ignore classroom interactions and instead aggregate data from different types of programs, operating under different sociocultural conditions and serving children with varying levels of first and second language abilities, are likely to be uninterpretable. In light of this warning, there is a need for a careful definition of the treatment and its implementation process at each replication site to aid the interpretation of observed effects.

(2) The need to describe and define the programs' participants and settings. Selection of replication sites was based on explicit suitability criteria that may pose a threat to the external validity of the evaluation design (Juárez and Associates, April 1978). That is, the ability to generalize results which emerge from the evaluation is related to the way in which the sites were selected. The site selection procedures specified by ACYF may yield a collection of sites that are atypical of those likely to be exposed to the particular curriculum models in the future. Therefore, it is necessary to carefully describe and define the program site's participants (i.e., children, parents, teachers, community, etc.) and setting so that appropriate generalizations of findings to similar well-defined groups can be made.

(3) The need to describe control group activities. The evaluation design calls for the random assignment of children to treatment (curriculum model) and control (non-Head Start) groups.

It should be clear, however, that the comparison of the treatment group is actually with the specific activities of the control group which have occurred during the time when the treatment group received the bilingual bicultural curriculum. If the control group activities are unspecified, it adds undesirable ambiguity to the interpretation of the effects. In other words, the outside classroom experience can become a plausible rival hypothesis in explaining differences. It therefore becomes important to accurately define and monitor the control group activities during the evaluation period.

(4) The need to observe classroom activities in various contexts. Test selection procedures were carefully developed in order to choose instruments that had a high degree of congruence with the curricular objectives to be evaluated (see Juárez and Associates, July 1978). Nevertheless, during the analysis of test items it became evident that it was not possible for the various objectives of the models to be assessed by the limited number of items in the instruments selected. In addition to this gap between test items and objectives, it also became obvious that while tests may provide some indication of change among the participants, they usually assess children's abilities in very restrictive contexts (i.e., the typical test situation). This observation is important, particularly in an evaluation involving very young children of various language abilities, considering the growing amount of evidence that much behavior should be viewed as an adaptation to particular tasks or situations (e.g., Cole and Scribner, 1974). That is to say, children possess a variety

of modes of functioning that respond to specific environmental demands. Therefore, a child's performance may depend on the nature of the social situation, including among other things familiarity with materials, previous experience with that particular environment, task expectations, other participants, and so forth. The implication of this position for the evaluation of young children is rather straightforward. If children are removed from a classroom situation in which they learned to use a particular skill (e.g., sorting) to be assessed in one that is less familiar (e.g., the test situation), competencies may be erroneously assessed. As Cole, Sharp, and Lave (1976) have suggested, performance is a result of an interaction between familiarity of content and familiarity of operations plus some knowledge of what constitutes adequate performance. Therefore, in addition to the pre- and post-testing, it seems important to assess children in conditions that match or parallel those conditions under which they learned the specific skills the programs are trying to teach.

Observing children under such conditions may also add to our understanding of the classroom dynamics which may have caused the observed effects. What has to be developed, then, is a strategy that goes beyond a reliance on test results as a measure of program effectiveness, and directly examines classroom activities during the course of the year (not just at the beginning and the end) in order to attempt to assess children's performance under conditions which resemble those situations in which they are learning the skills the programs teach, while also attempting to

define which educational activities promote the desired program outcomes.

In the next section the specific objectives of such a strategy will be discussed.

Objectives

During the initial phase of the evaluation we suggested to the ACYF staff members and the curriculum model developers that an important evaluation task should be to acquire multiple measures of the program's process and outcome behaviors (see Juárez and Associates, 1977). This strategy was intended to explore the competencies and abilities of the program's principal participants, the children, through a series of logically related observations and other measures in naturally occurring classroom contexts. In general terms, the goal of such an approach was to avoid exclusive reliance on pre- and post-testing as a measure of very young children's competencies (or program effectiveness), and instead carry out observations of their behaviors over time and from multiple perspectives to assess the effects of the model programs.

In this report, therefore, as in previous documents (Juárez and Associates, April and July, 1978), we seek to further elaborate on and develop the notion that in order to achieve the contracted evaluation goals, the experimental psychometric approach should be complemented by a multi-method data collection strategy. As we see it, there are three primary objectives to such a strategy. They can be summarized as follows:

(1) To assess the nature and extent of program implementation over time. The implementation evaluation will seek to provide information as to the feasibility of successfully implementing the models in more than one setting. This information would entail assessing the congruence between the proposed programs as written by the curriculum model developers and the programs actually implemented at the replication sites. Information will also be collected on the degree to which the various programs were implemented at each site.

(2) To use information gathered from observations to help interpret the quantitative (test) results. Observations of classroom activities and situations, such as specific lessons, and language use in various situations, will be taken into consideration in the interpretation of test data. These observations will also include information on the measurement process or how the children in both the treatment and the control groups were tested, and on the experimental arrangement, including information on the control group experience and how well randomization worked. All of the above activities may have a strong influence on test results

(3) To provide quantitative change-data over time. Focusing on model-relevant child behaviors, observational data will be systematically collected during the evaluation year. This procedure will provide a complementary and an alternative way of examining the program's impact on parents, teachers, and children. It will enable the observation of program effects both on domains specified by the model developers as well as on behaviors not anticipated

by the model developers. It would also provide a methodologically independent cross-validation of the quantitative test results (see Campbell, 1974).

In our view, to carry out a multi-method, contextual evaluation study that combines both quantitative and qualitative methodologies, it is essential to carefully train and place at each evaluation site a "participant researcher" (PR). These on-site evaluation staff members will supervise the psychometric data-gathering and will become the primary people responsible for collecting data on both the school and the control group (Juárez and Associates, April 1978).

Procedures

The Participant Researcher

There are several reasons why Juárez and Associates advocates such a participant researcher role. It is our position that too many evaluations involve sending in outsiders who do tests, relatively infrequent interviews, or questionnaire measures, and then leave. A common complaint is that this type of evaluator does not ask questions relevant to the programs on the sites, and fails to get a fair and balanced sample of the typical life within the school. Such evaluators cannot really know what is going on, or know nearly as much, in some ways, as the insiders who run the school. A PR is both an insider and an outsider; his/her role helps to overcome this structural problem of differentiating between the roles of "outsider" evaluator and "insider" staff member

of the education program, and his/her participation helps provide contextual and relevant data at each site.

Research Strategies

The PR, as mentioned earlier, will use several data-gathering methods. The specific data collection procedures are summarized below:

(1) Naturalistic observations. These observations will include event samples of specific participants' behaviors in context, observer ratings of specific evaluation domains (e.g., socioemotional development), and focused ethnographic notes on program implementation and the general functioning of the site.

(2) Teacher interviews. Both structured and unstructured teacher interviews will be collected on implementation, attitudes, opinions, etc.

(3) Phone-call updates. These phone calls will facilitate the collection of daily-routine information of both treatment and control group parents.

(4) Psychometric testing. The PR will also supervise and coordinate the pre- and post-testing activities specified in the contract.

All of the activities summarized above are discussed in more detail in the section on Evaluation Procedures in this document.

A Preview of the Discussion

In the following pages we will present a detailed plan for the use of observational and other strategies as necessary

supplements to the experimental design. This preliminary document will proceed by discussing the characteristics of the approach adopted to facilitate the collection of both observational and test data. The discussion then concentrates on the information needed to satisfy the need for assessing program implementation and impact. For the latter category, this includes classroom behaviors and control group activities.

Subsequently, a strategy concerning needed personnel and a plan for conducting the evaluation are presented. The function of this plan is to describe the role of the participant researchers, including how these individuals will be recruited, trained, placed at the evaluation sites, and supervised. Following this, data collection activities of the participant researcher will be discussed and examples provided. A section on data management strategies and the monitoring of data quality is also included. A plan for data analysis and interpretation concludes this report.

II. Evaluation Procedures

A. A Change Over Time Approach

The evaluation approach developed by Juárez and Associates attempts to integrate, as part of an experimental design, qualitative and quantitative methodolgies as complementary data-collection strategies. That is, the mix of both a qualitative and a quantitative evaluation within the framework of an experimental design is seen as feasible and desirable. Campbell (1974), for example, has pointed out that "the equivocalities of inference which experimental design seeks to reduce are present for qualitative evaluation fully as much as for quantitative" (p. 18-19).

In order to facilitate the integration of both methodologies, a time-series approach to data collection has been adopted. This arrangement permits the administration of pre- and post-tests and it also facilitates periodic and systematic ongoing classroom observations, while allowing the field worker sufficient time for analysis and writing as for the actual field work. Basically, such an approach involves the testing of program participants at the beginning and conclusion of the treatment, plus classroom observation for four specified and preselected two-week periods (see Appendix A for approximate dates). These naturalistic classroom observations focus on samples of approximately 15 children per site. This sample can be stratified by, for example, the children's language preference, sex, and/or age. Limiting the observations to a sample of children instead of

observing every participant makes a difficult task manageable and facilitates the use of multiple, but complementary, data-gathering techniques.

Other important characteristics of this approach can be summarized as follows:

1. Field workers are familiar with and informed about the evaluation setting (non-naive).
2. Important program or site-specific variables not shared by the other models are assessed. These include exogenous variables that are observed to have an influence on program effectiveness.
3. The assessment program implementation process is emphasized.

Both the quantitative and the qualitative data-collection efforts can focus on the same dependent variables.

These overlapping forms of study require that the evaluators limit and specify the program objectives to be assessed. What follows, then, is a discussion of the variables that compose the two major categories to be studied: program implementation and program impact.

Study Variables

Program Implementation

The goals of the evaluation segment concerned with program implementation are to provide information to (1) parents, teachers, and administrators who have an interest in initiating a Head Start program based on one or more of the models, and (2) individuals who are concerned with the impact of the models on children, parents, and teachers and seek to learn the extent to which a respective model was actually put into operation relative to the way it was described by the model developers. Implementation evaluation will attend to both individual and cross-model concerns and will involve observations and interviews. (These methods are to be discussed in a subsequent section.) The implementation evaluation conducted by Juárez and Associates is not intended to substitute for any implementation assessments carried out by the curriculum model developers. Instead, the Juárez evaluation is intended to provide the basis for the preparation of pamphlets describing the model programs as well as the basis for a better understanding of the impact of the individual and combined programs on parents, teachers, and children.

This document is based on the assumption that implementation is an ongoing process requiring an assessment strategy that parallels the program activities and continues throughout the entire experimental period. The variables of concern to the developers of curriculum models and the evaluation contractors

in this assessment process are likely to be both overlapping and unique. The curriculum developers, for example, may be concerned with the formative evaluation of teacher training procedures or the instructional materials they prepared to enhance the effectiveness of the models themselves. The evaluation contractor, on the other hand, may be concerned with whether and to what extent the model is implemented as well as with assisting others in learning about the curriculum models. The conduct of implementation evaluations by both parties should enable these related objectives to be met.

In each instance, the implementation evaluation will seek to provide information as to the feasibility of successfully implementing the models in more than one setting. It will also provide information to others regarding both the process necessary to implement each model and the procedures needed to maintain each model in operation. Such knowledge should help others to learn of the ease of implementing each model in a new environment, including a description of any special personnel, parents, students, or community needed to assure a greater probability of success in the implementation of the model. In accord with the evaluation contract, the extent to which a model has been implemented will be taken into consideration in the interpretation of impact data.

In evaluating the nature and extent to which each of the curriculum models is implemented, there are five proposed categories of information to be collected. These include (1) the setting and material resources available; (2) the individuals

involved; (3) the schedule and organization of the programs; (4) the attitudes of the individuals involved; and (5) the activities in which the individuals are engaged. Because the major purpose of this document and of the preliminary discussions associated with it in the Washington mid-December meeting is to seek the advice and counsel of the curriculum model developers on the collection of evaluation data, it is important to identify the way these five categories of information were chosen. First, a review of the programs and objectives of each of the curriculum models as well as a review of each model's plan for assessing its own implementation were undertaken. This pool of inductively generated items was then combined with variables identified in other evaluation studies. The combined list of variables are reduced and organized thematically into the five categories mentioned above. The principal concern throughout was to assure that the items included would be relevant to the models while enabling Juárez and Associates to be responsive to its own ACYF contract objectives. These five content areas were also deemed important in the overall evaluation, since they form part of two ACYF contracted relationships being assessed.

The first relationship concerns the congruence between the proposed programs as written by the curriculum model developers and the programs actually implemented at the replication sites: in other words, the relationship between words and deeds or that which is intended and that which is manifested. Knowledge about the site, personnel, schedule, activities, and so on will provide

a basis for knowing something about this relationship and will provide insight into the program as funded as opposed to the program as implemented. Without such information, it will be difficult or impossible to inform others as to the nature, content, personnel, and so on of the program except as it was originally proposed. This is especially crucial to the dissemination of information to teachers, parents, and others who subsequently seek to implement one of the models in their own setting.

The second relationship concerns the degree to which the various programs have been implemented. Here the issue is not whether the implemented program has any relationship to the proposed program but instead the extent to which the program was implemented at all. Without this information the impact data may be interpreted as if the program had been fully implemented or implemented as proposed. In reality, however, there are likely to exist many constraints and obstacles to fully implementing any given model, causing certain compromises to be made in the operation of the program. For example, the site may be inappropriate or poorly equipped, the personnel may be poorly trained, the community may not desire the program, instructional materials may not arrive on time, and so on. Without some knowledge of these constraints and their influence on implementing the program as intended, the impact may be unjustly interpreted or interpreted as if the program were unencumbered.

The data needed to address the five categories of information to be collected and specific questions related to each are

detailed below.

The Setting and Material Resources

This category concerns the relationship between the physical environment, including equipment and supplies, and the program needs and goals. The program models vary somewhat in the attention paid to this category with, for example, one seeking considerable open space and another seeking more structured space. Whatever the need, however, it is assumed that the availability of an appropriately equipped and supplied site will affect the nature and extent to which a program can be implemented. It is further assumed that this site suitability will influence the interpretation of the program's impact and the decision of other interested persons as to the transferability of the program to other locales.

1. Do the space, equipment, and supplies reflect the program assumptions and goals?
2. To what extent does the site meet the criteria established for buildings, equipment, and materials?
3. Can various areas (e.g., blocks), equipment (e.g., tables), and supplies (e.g., books) be identified on a map of the area?

Individuals

Here we are concerned with all of the individuals--ACYF representatives, curriculum model developers, administrators, teachers, aides, parents, and community resources--who will be involved in implementing the program at a given site. It is

important that certain demographic data relative to these individuals as well as their participation in the program be documented. Furthermore, we will want to know whether the children and their assignment reflect the criteria as established for the implementation of the models. This information will be of special use in interpreting the impact data and informing others of the individuals needed to ensure a successful program.

1. Describe the administrators, teachers, aides, and community resources involved in implementing the program.
2. How often do curriculum model representatives and ACYF representatives visit the site and what is their routine and activity on each visit?
3. To what extent does the site meet the criteria for the number and composition of children by language ability and cultural background?
4. To what extent are individuals other than parents, teachers, administrators, aides, and children in the classroom (e.g., siblings of the enrolled children, community members, visitors)?

Schedule and Organization of the Program

This third category of information documents the extent to which planning is evident in the program, the nature of the record-keeping that occursin the center, the decision-making and the decision-making structure, the grouping of children and others, communication channels, and so on. In each case, we are interested in how the program is structured and organized to achieve the stated goals. This information will be especially

important to other individuals who seek to learn about how to initiate a similar program in their own locale.

1. Do the program activities follow a predetermined plan (e.g., daily lesson plans; weekly plans; unit plans)?
2. What criteria are used to alter these plans?
3. To what extent do parents, teachers, and aides use information collected about the children (e.g., attendance, emotional changes, cognitive changes) and the children's progress for purposes of enhancing the children's learning?
4. To what extent do children work alone or in groups of two or more? Teachers? Aides? Parents?
5. If children, teachers, aides, or parents are grouped, what criteria are used to establish the groups?
6. To what extent do teachers, aides, and parents work alongside and with children?
7. When and in what ways are parents and community members informed of events and meetings associated with the center (e.g., telephone, newsletter, face-to-face)?
8. Under what circumstances and how often do administrators, teachers, aides, and parents meet formally or informally?
9. How are decisions made at the site (e.g., who participates, what is the structure)?
10. To what extent does the site meet the criteria established for randomly assigning the children to experimental and control groups?

Attitudes

To some extent the support of the administrators, teachers, aides, parents, and the community at large will be needed to ensure a successful program implementation. A few appropriately placed individuals who are outspoken in their negative or positive

appraisal of a program may well alter the outcomes. This category of information seeks to identify attitudes of this sort as well as the expectations and explanations held about the program, interpersonal relationships, and the history of the center as it impacts on the present effort. This information will assist in interpreting the impact data, especially in those instances where conflict or negative appraisals predominate.

1. What arethe opinion of teachers, aides, parents, and community members about the program and its operation?
2. Does the model developer's expectations about the program's direction and emphasis remain consistent or does it undergo change? What about the expectations of teachers, administrators, aides, and parents?
3. To what extent are admininstrators, teachers, aides, and parents able to articulate and explain the relationships among the model's assumptions, activities, and goals?
4. To what extent is the social climate of the program characterized by interpersonal harmony or conflict?
5. To what extent do administrators, teachers, aides, parents, and curriculum model developers believe the model has been implemented in accord with stated intentions?
6. What is the history of this particular center?

Activities

The final category of information related to implementation concerns the activities or what happens in the centers. Here it will be possible to test the relationship between what teachers,

children, and parents actually do and what the curriculum developers hoped they would do. Likewise, information as to typical routines, the emphasis placed on language, cognition and culture, and so on, will be acquired. These data will provide some insight as to the relationship between activity and learning or activity and outcomes as well as providing interested individuals with information about what actually occurs in a setting where a particular model is implemented.

1. To what extent do the activities of children, parents, and teachers reflect program goals and assumptions?
2. To what extent do children explore the setting as opposed to being directed to participate in an already established set of activities?
3. To what extent can a representative sample of daily, weekly, and monthly routines be described?
4. To what extent do the activities reflect a bicultural or multicultural emphasis (e.g., print media, film, audio records, displays, special occasions)?
5. To what extent do the activities provide an opportunity to acquire a second language as well as to reinforce a primary language?
6. To what extent and in which activities are children provided the opportunity to learn in their preferred cognitive style? To become bicognitive (Santa Cruz model only)?
7. To what extent and in what ways do teachers, parents, and aides receive pre-service and in-service training?
8. In what ways does the program monitor the health status of participants (e.g., vision, hearing, nutrition)?

The information needed to assess program impact is discussed

below. This category has been subdivided into classroom behaviors and control group activities. As in the previous section, some of the procedures used to identify relevant variables will be identified and discussed in the text.

Program Impact

Classroom Behaviors

As mentioned earlier, one of the characteristics of a focused and systematic qualitative data-collection strategy requires knowing something about the setting one is studying through information gathered before entering the setting (Erickson, 1977). For the purposes of the present evaluation, the assessment of classroom behaviors first involves some prior knowledge of the orientation, characteristics, and objectives of the four bilingual curriculum models under study. Building from a list of objectives supplied by the program developers, it is possible to specify several program-relevant behavioral domains and components. Thus, for example, a contract specific domain (e.g., concept development) is first reviewed theoretically, resulting in its tentative breakdown into separate components (e.g., seriation, classification, etc.). These components are then assessed against the objectives of the four curriculum models to determine if, in fact, these components were going to be taught in the programs. Once the components have been identified, they are divided into several participant behaviors that can be

observed empirically (e.g., the child arranges objects in sequence). These components and their respective behaviors are purposely limited in number so that they can be studied in greater depth and from a number of perspectives. This reduced list of components and their related participant behaviors were then reviewed by program representatives for accuracy and completeness. At the conclusion of these initial procedures we had a manageable and accurate list of across-model objectives that can be assessed by both tests and observations (see Appendix B).

The next step in carrying out this type of evaluation is to specify both the setting and the context in which the behaviors are to be measured or observed (for example, see Appendix C). "Setting" refers to the macro environment, such as the school or playground. Within each component, behaviors are identified according to setting as taking place in school or in other locations. "Context" refers to a particular micro environment which includes the specific physical setting (e.g., reading rug, block corner, play swings), the participants (e.g., two children and two aides), what they are doing, and when they are doing it (Erickson & Schultz, 1977). The identification of observable behaviors by setting and context is crucial given the growing concerns over the complex influences of the environment on individual action (see, for instance, Cole, Gay, Glick & Sharp, 1971; Cole & Scribner, 1974) and our position that evaluation data should be interpreted in relation to the situation in which they were gathered.

The importance of taking into account the social situation in which data are gathered in an educational evaluation concerns the relationship between the learning environment and the assessment environment. The learning environment refers to the conditions under which the experiences are introduced (e.g. sorting) and become part of an individual's behavioral repertoire in that specific context. The assessment environment refers to the conditions under which an evaluation or behavioral assessment is made. We assume that the closer the relationship or match between these two environments, the greater the prospects of securing valid performance data in the evaluation. First, we expect that the social situation (e.g., stimuli operations, expectations) will be a familiar one to the children, thereby attending to three important factors that have been suggested to influence performance (Cole, Sharp & Lave, 1976).

Second, we believe that by attempting to carefully choose the conditions under which performance is assessed such that they reflect the influence of the learning experience, it is possible to overcome some of the problems associated with learning transfer in young children (e.g., Brown & Campione, 1978). That is, we view the educational process at school as providing learning experiences for specific performances. Hence, we would look for performance evidence (i.e., assessment) under conditions which resemble as much as possible the learning conditions of the school.

Control Group Activities

The control group will be followed through a telephone updating procedure during the course of the evaluation year, between the pre- and post-test testing interviews. A randomly selected subsample of fifteen families will be called approximately once a month and information will be gathered on a number of topics directly related to home language use; exposure to other educational and preschool experience; change, mobility, and variability in residence and household personnel; and child health.

The fundamental reason for collecting this information is our knowledge that the control group families will not be idly spending their year waiting for the evaluation to end. It is very plausible that many of the families not selected for the Head Start Center will seek out alternative preschool, daycare, or other educational experiences for their children. They may differentially expose their children to educational television, to English-language environments and caretakers, and so forth. We feel very strongly that it is important to monitor such data during the year. When the comparison and control group families are compared to the Head Start group after the post-tests are done, Juárez and Associates will then be in a far better position to account for, interpret, and understand whatever similarities or differences may emerge between the two populations. Thus, these data are being collected to aid in understanding and interpreting results, rather than as a measure of outcome. If a given model shows no significant differences between control

and Head Start children on second language comprehension for example, and we have data only on the treatment within the Head Start group, the only information available for interpreting this finding is a retrospective examination of the treatment. However, it may be true that the comparison group put their children in other preschool experiences, or had other language exposure similar to that experienced in the Head Start Center. In short, we are collecting control group information during the year to meet one of the aims of the original contract. This aim is to understand the strengths, weaknesses, and contributions of each model.

Telephone call updates will focus on topics related to educational development and language use. These general topics will include:

- Any other preschool, daycare, Head Start or non-parental caretaking arrangements which have been experienced by the control group sample.
- Characteristics of preschool, daycare, etc., and experiences of the children.
- General information on the typical weekday schedule and activities of the control group children including what major kinds of settings they are in, who their major caretakers are, and their language facility as judged by the parent or the primary caretaker.
- Variability and change in the residential, home, and family life of the control group children during the course of the study. These will include residential changes, changes in household and domestic group personnel, role of male adults (fathers, semi-absent fathers in the home), school and educational experience

. of older siblings or co-residents in the household, visits to other parts of the country or to Puerto Rico, Mexico, and so forth.

Overall Strategy: The Participant Researcher Role

We propose that for a year a full-time on-site "participant researcher" (PR) or field worker be placed in each Head Start site being evaluated. These on-site evaluation staff members will be the primary people responsible for collecting data on the school and control groups. This participant researcher role is an innovation in evaluation studies, and we believe it to be relevant and appropriate given the differences in the four bilingual Head Start models that will be evaluated.

Too many evaluations send in outsiders who do tests, relatively infrequent interviews, or questionnaire measures, and then leave. A common complaint is that this type of evaluator does not necessarily address the questions at each relevant site, and that the data collection procedures are not conducive to obtaining a balanced sample of all of the behaviors or of the typical life within the school or classroom. They do not, in other words, really know what is going on or know nearly as much, in some ways, as the insiders who run the school.

A PR or insider-outsider evaluation strategy would help overcome this structural problem in the roles of evaluators and full-time Head Start staff, and would provide contextual and relevant data at each site. The PR primary role definition and responsibilities involve collecting both qualitative and

quantitative evaluation data. Thus, the PRs have an important and difficult role: for selected periods of time they are simultaneously involved in the day-to-day activities of each Head Start center and know the inside, or what is "really" going on--what teachers and other staff members think about the school, how parents and children react on a day-to-day basis, etc. But PRs are also responsible for the coordination or supervision of the collection of formal test measures, for interviewing the parents and teachers, and for contacting a sample of homes to gather ongoing information on educationally relevant activities.

Recruitment

Juárez and Associates would like to start recruiting candidates for the PR as soon as possible (January-February 1979), since the type of staff member we envisage might need quite a bit of lead time in order to prepare to take the job. A major recruitment effort will be undertaken through notices in the newsletters of relevant professional organizations, and through contacts with university departments of, for example, sociology, anthropology, education, as well as with research training programs and experienced field workers. It is recommended that the selection of the PR should be the joint responsibility of Juárez and Associates and each curriculum model. Since the primary characteristic of the PR is the ability to work effectively in the field with the Head Start centers, the recommendations of the Head Start site staff will also be carefully considered. However, the final selection

of the candidate will be the responsibility of Juárez and Associates

Job Description

Ideally, the PRs must have knowledge of the research role and issues involved in data collection. As employees of Juárez and Associates they must understand their roles as evaluators; they must be sensitive to local circumstances, and be able to work effectively in the Head Start center for the year. The search for a balance between these two important needs will be an important factor in the selection of the candidates.

It is preferred that the PR be at or around the M.A. level in some social science or educationally related discipline. The PR must be bilingual, would start on or about July 1, 1979, and would commit him/herself to one year of full-time field work and data collection. We recommend a starting salary for such a person in the range of $14,000 or higher. It is anticipated that previous experience, knowledge of the Head Start site, field acceptability and writing skills will substantially influence the choice of PRs and will be strong supplemental criteria in addition to formal education.

The important points in the job description for the PR, then, include the following (see Appendix D for a summary):

(1) At or around the M.A. level

(2) Experience with early childhood programs in addition to formal education

(3) Bilingual

(4) Willing and able to make a commitment to work for a year

(5) Available on a full-time basis

(6) Able to work effectively with the models and the Head Start center staff

Training

The training of PRs will be a mixture of on-site experience and involvement with the Head Start centers, training in the coordination of formal testing procedures, and classroom observation and interview procedures.

An initial training session will be conducted in Los Angeles by Juárez and Associates in the summer of 1979. This initial training period will be important for developing consistency in data collection techniques and consistency of overall planning and conceptualization of the project solidarity among the PRs.

The training of the PR will rely on an extensive program for the training of naturalistic field observers developed over the past six years at UCLA (Weisner, Turner, Gallimore, and Levine, Laboratory for Naturalistic Observation). This training program emphasizes most of the major problems field researchers face and offers specific techniques for the implementation of effective observational strategies for solving these problems in the field. The topics covered in these courses include the following:

1. Initial immersion and exposure to field work problems.
2. Three different techniques for writing effective narratives of field observational data.

3. Techniques for managing one's role as a PR in the field setting, including techniques for disclosure rapport, informed consent, and ethical issues.
4. Techniques for coding and systematically obtaining field notes based on observations, including surveys, different coding schemes, post-coding, and rewriting field notes.
5. Techniques for both formal and informal interviews, including questionnaires, focused interviews, informal question-asking, and "surreptitious" informant interviewing.
6. Systematic and focused field observation, including nine different techniques for doing field observations, such as time samples, event samples, key incident samples, and various kinds of qualitative accounts.
7. Techniques for keeping a personal diary and record of the sequence of events and the key events at the field site.
8. Methods for data reduction and data analysis of qualitative and quantitative materials collected in the course of PR observational field research, including combining quantitative and qualitative data, analyzing events across time, and systematically attempting to measure and control for reactivity and observer bias.

Specific field exercises have been designed for the introduction of each of these problems and will be adapted to the needs of the PRs and each of the sites.

The most important method for training the PR will be practice in actual observation at local Head Start sites, in addition to some classroom work and readings in observational methods. The PR will go to Head Start centers in the Los Angeles area and carry outa variety of pilot and practice observations of the children and the teachers. Similar direct field practice will be utilized in doing interviews with teachers and with parents. Direct experience with field observation, actual modeling of the appropriate kinds of techniques, and discussion of these techniques in the field are by far the most effective training methods. Readings and other materials will supplement this direct field experience. A bibliography of more than 50 items, most of which are annotated, is available on studies in naturalistic observation and can be utilized as part of the training program. However, direct field experience and training in specific observations and field work to be done at each site are the core of the training process.

These training procedures will be supplemented with periodic visits by a "model" or "standard" PR coordinator to the field sites to do parallel on-site observations in the classroom to ensure that reliability of data is adequate and that there has not been excessive drift away from the initial standardized procedures. Additonal staff meetings of PRs at Juárez and Associates in

Los Angeles for debriefing will take place during the course of the evaluation and at the conclusion of the data collection in 1980 (see Appendix A).

Placement

As the Gantt chart in Appendix A indicates, the participant researchers will be on-site for initial contacts with the Head Start program, the model developers, and the community by Jul 1979. This period of initial establishment of the participant researcher's role and field status and of simply getting to know the people s/he will be working with throughout the year, is essential to the functioning of the participant researcher and to the data collection. The participant researcher will need time initially to explain what s/he doing, to do some pilot work on-site, to make the initial contact with the treatment Head Start and control groups of parents as well as with the relevant committees and school personnel in the area, and to set up housekeeping.

The participant researcher also needs to be on-site early to help coordinate testing. The participant researcher may be doing some of the testing him/herself but will also be assisting and coordinating outside testers at the start of the year. Effective coordination implies that the participant researcher is well enough established to be able to perform this role.

The PRs first impressions of staff at the Head Start

site and of the programs in general are also very important. It is a well-known phenomenon in field research that detailed attention to the earliest impressions are vital later in the study. Obviously many of the first impressions are uninformed and may be wrong or biased, needing to be put into a context that only time will provide. However, many of the things about the site that will be observed early on and that will jump out at the participant researcher early in the year will be vital and fres information that should be carefully recorded for comparison with the final assessment and write-up of notes at the end of the project.

Coordinator

Juárez and Associates does not underestimate the difficulties involved in managing participant researchers in the field and collecting information in order to make an adequate evaluation. For this reason, there will need to be a full-time coordinator/supervisor for the PR-related aspects of the evaluation and data collection efforts. The project director is responsible for selecting and assigning a coordinator to this task. Although this person will be based in Los Angeles. s/he will spend a fair amount of time in the field coordinating the data collection efforts at the various Head Start sites.

The coordinator's role is a critical one in assuring the success of the PR strategy. There are obvious difficulties related to timing and local needs that will have to be weighed

constantly. There will no doubt be some changes in the PR's role during the period of evaluation, and there will need to be consultations and scheduling of the various data collection efforts. The coordinator will be responsible for setting deadlines for and directly supervising the PR, for monitoring the reliability of field observations and other data collections done by the PRs, and for ensuring that the timing of data collection efforts meets the needs of each model site and that the roles of the PRs are appropriate to their settings. S/he will also have principal responsibility, along with the project director, for hiring the PRs. The coordinator will also be making calls to the field him/herself to spot-check telephone interviews and other contacts with the families and the schools for comparison and reliability purposes.

In short, the PR strategy will work if the day-to-day managerial tasks are carefully and systematically handled. The PRs will need supervision and direction, and the strategy will not work in a coordinated way across all the sites unless the full-time coordinator (assigned and supervised by the project director) is available. A summary of the preferred background characteristics for such a position is included in Appendix E.

Data Collection Activities of the Participant Researcher

There are a variety of possibilities for the structure of the PR role. Juarez and Associates believes strongly that those collecting data in the field should primarily be evaluators: they should have some understanding of field or experimental research traditions; and their primary role definition and responsibility should be to collect evaluation data, as mentioned earlier. The amount of data that PRs will be asked to collect and the level of skill they must have, necessitate that they have training in social science-related fields and make it preferable that they have some experience as researchers.

The participant researcher will utilize a variety of data collection techniques in the field. It should be clear that all of these methods overlap and can apply to the assessment of program effectiveness of impact. Thus, the duties of the participant researcher while in the school setting will be a mixture of (1) direct observations of a subset of Head Start children and all teachers in every classroom, (2) ratings, and (3) qualitative notes focused on evaluation domains and program implementations. The following descriptions and examples will clarify the nature of these activities.

(1) Focused systematic classroom observations. An instance of a currently planned area of observation is the children's use of language. Time and event samples of children in the classroom using direct and focused naturalistic observational methods will be employed to monitor English and Spanish language use in various contexts.

Examples

Figure 1 shows a hypothetical event sample record sheet for recording data in the domain of Spanish-English language usage for Head Start children during the classroom period. The "events" column simply lists a series of specific behaviors related to language use. The descriptive categories underlined (conversation, fantasy/meanings/pictures, complexity, switching, etc.) all refer to general components concerning language use which are being measured for the contract.

The participant researcher selects a target child at successive predetermined time intervals (5-15 minutes per child per sequence, for example) during the morning period, focuses on that child, and records various kinds of language the child uses. Using the subsample of 15 Head Start children in each classroom, such information could be collected two or three times a week for each child in a variety of different contexts.

The "code" column gives some examples of how an event sample record would be used. If Maria describes something about the class to José (a teacher) in Spanish during freeplay time outside, this would be coded as per the example in Figure 1. (Maria-José T-S-2; the notes accompanying Figure 1 give more details). The point of this hypothetical events sample record is to emphasize the language use data that can be collected for each target child (or teacher) taking into account a variety of different contexts and types of language use.

These data can be analyzed in a number of ways. For example,

the frequency of Spanish-English or mixed language use can simply be aggregated and summarized for each child, for each context domain of language, or for various settings or dyads (such as a particular teacher to all children, or same-sex to opposite-sex children, etc.). This information can also be aggregated by diversity of events recorded for a child, diversity of language use for a child, whether the child initiates or is the recipient of English or Spanish language from others, and the context in which a given child or given dyad most often appears. For example, is most English spoken in classroom activities, or is most Spanish spoken between children outside during free play. The reason for collecting data on target children in a relatively disaggregated form is so that summary measures across many different observational sessions for the target children can be compiled and aggregated in different ways by the participant researcher. At the end of the in-class Head Start observation period, the participant researcher will compile such summary statistics for submission to Juárez and Associates.

It should be emphasized at this point that this is a purely hypothetical recording sheet, and that the events listed are not necessarily meant to appear together on the same event sample record. Obviously many of the events overlap. For example, a child could give instructions to another child using compound sentences in both Spanish and English. Not all of these would be coded by the observer simultaneously. Rather, the observer would have one event sample sheet measuring language complexity

Figure 1

Hypothetical Event Sample Recording Sheet for Spanish/ English Language Usage of Children[1]

ents[2]	Code[3] (Examples)
scription	
scribes common objects	
escribes class events	Maria-Jose-T-S-2/[4]
alks about self	
alks about others	
alks about fe lings	
onversation	
ives instructions	Janis-Luis-E-4/[5]
ells a story or event	
antasy/Meanings/Pictures	
okes, laughs over stories, pictures	
atches words with pictures for child	
omplexity	
roduces complete sentence - simple	Tom-Ray-M-1/[6]
" " - compound	
" " - complex	
witching	
ses both Spanish & English words in sentence	
eterminates in S. & E. sentences or phases	
nswers in S. or E. when addressed in S. or E.	

Notes for Figure 1

1. Completed by participant researcher in classroom settings at standardized periods during the morning, focused on specific teachers, or children, in a timed sequence.

2. Events will vary by domains in contract and by key events in each model's implementation list. This sample list includes several different components for one domain (Spanish/English language production), and overlap in content for illustration.

3. Each event for a focused observation of a child or teacher is coded using a simple mnemonic such as the following examples.

4. Maria (target child) describes a class event to Jose (teacher) (T) in Spanish during free play time outside.

5. Janis (target child) gives instructions to Luis (other child) in English during an object manipulation class activity.

6. Tom (target child) talks in mixed English and Spanish to Ray (other child) using complete sentence.

Sample Code for Figure 1

Language Code:	S-Spanish E-English M-Mixed
Teacher Code:	(names)
Child Code:	(names); or M/F and age
Context Code:	1 story-telling time 2 free play outside 3 free play inside 4 object manip. activity 5 word/picture 6 other

and switching, for example, and another for kinds of description and conversation produced by the child. The event sample recording sheet illustrates one method of collecting and quantifying language use scores that are context and domain specific and that could be compared across the full year.

As stated earlier, these records, although designed to collect teacher and/or child language data, could also be used to assess implementation. For instance, a model may specify that during a classroom activity time teachers should use a mixture of Spanish and English and should encourage the children to do the same. These event sample records could be used to answer the question, Was a mixed language strategy employed by teachers and produced by children in this particular context?

More simplified event sample recording is, of course, also possible. For example, the participant researcher may elect to focus on a particularly important and telling classroom context, such as, for example, a period of 20 to 30 minutes each morning when the teacher and children tell stories. Language use may be particularly rich during this period. The participant researcher could utilize a context-specific event sample record just for this period of time. Without focusing on particular children, teachers, or dyads, the events list in Figure 1 could be adapted to a checklist format. For example, each time teachers or children talked about feelings, the participant researcher would simply put a check in the corresponding row. Each time Spanish was used by a child to talk about feelings,

that row would be checked and the language specified. In other words, a pure frequency count of all of the language-related variables of interest during that particular setting-context--e.g., story-telling time--would be counted up. This would give a comparative measure across the year of the frequency of Spanish and English usage, the way language is used by teachers and children as a group, and the content of the activity. If a given model pays particular attention to such a key context or key events, the participant researcher may find such a context-based frequency counted checklist to be most effective.

These kinds of measures will be repeated across children and across contexts during the entire year. Thus, changes and increases or decreases in the behaviors in context will be measured. Also the stability of the measures and their validity will be enhanced, since no child or context will be judged using only one or two visits. There will be repeated observations of the classroom throughout the year, providing a comprehensive and productive measure of what is happening within the school.

(2) Participant researcher ratings. Direct classroom observations are a powerful but relatively difficult and complex method of data collection. Direct observations will be supplemented with a weekly observer of, for instance, certain standard dimensions of language use (illustrated in the previous example) for each site.

Examples

Figure 2 gives a hypothetical example of some rating procedures which a participant researcher could use in assessing language skills. The dimensions and the scale vary depending on the focus of the variable being measured. The language usage is measured relative to specific dyads (teacher-child; aide-child; other children to children of the same sex; other children to children of the opposite sex; the child talking to himself; and an overall judgment). This dyad-specificity seems extremely important as a way to specify the context within which children are more or less likely to use English, Spanish, or mixed language.

Each target child will be rated at the end of each week of participant-research work within the classroom. Although the participant researcher will be involved in a variety of classroom research activities, one of his/her major goals will be to focus on the children and on language use in the classroom well enough to be able to make such ratings whenever possible. (Obviously if the child was absent or there was no opportunity to observe, certain ratings may be missing for certain dyads.) On the whole, however, a participant researcher in the classroom on a daily basis for ten weeks over the school year should be able to make a series of valid and informed judgment about language use. These

Figure 2

Hypothetical Example of Rating Procedure for Participant,[1] Researcher Use in Assessing Language Skills of Head Start Children

Child _____ Week _____ Rater _____

Pgm/site _____ # days child attended _____

CHILD[2] LANGUAGE USE - ENGLISH[3]

a. Use the following scale to rate the child's talkativeness in each of the dyad situations listed.

Scale:

0 Not applicable
1 Rarely spoke, quiet
2. Fairly quiet, some talk
3 Some talking
4 Talked often
5 Very talkative, voluble
9 Missing

Dyads:[4]

_____ Teacher-child
_____ Mother-child
_____ Aide-child
_____ Other children-child (same sex)
_____ Other children-child (opposite sex)
_____ Self (child)
_____ Overall (child)

b. Use the following scale to rate the complexity and sophistication of the child's utterances in each of the dyad situations listed.

Scale:

0 Not applicable
1 Non-linguistic utterance/or single words primarily
2. Mostly two or three word short sentences
3 Frequent use of short sentences (noun-verb modifiers)
4 Frequent use of long, complex sentence construction (noun-verb modifiers-prepositional modifiers-connectives)
5 Frequent use of long paragraph speech patterns composed of complex sentences

Dyads:

_____ Teacher
_____ Mother-child
_____ Aide-child
_____ Other children-child (same sex)
_____ Other children-child (opposite sex)
_____ Self (child)
_____ Overall (child)

Figure 2 (cont'd)

c. Use the following scale to rate the frequency with which the child engaged in dialogue with another person (any sequence of verbal interaction between the child and another). Circle the person who most frequently initiated the dialogue.

Scale:

[illegible] Not applicable
[illegible] Not at all or rarely
2 A little
3 Sometimes
4 Talked often
5 Very talkative, voluble
9 Missing

Dyads:

_____ Teacher
_____ Mother-child
_____ Aide-child
_____ Other children-child (same sex)
_____ Other children-child (opposite sex)
_____ Self (child)
_____ Overall (child)

d. How socially appropriate is the child's language use?

_____ 1. Inappropriate: Child talks without regard to people present and/or does not acknowledge another's speech much of the time.
_____ 2. Inappropriate only some of the time.
_____ 3. Appropriate: Child talks when speech clearly required. Acknowledges content of another's speech. Initiates speech only when relevant or totally appropriate.
_____ 4. Appropriate and somewhat enriched. Not only is the child's talk appropriate, but sometimes the child enriches a social conversation with talk that is not necessary but socially pleasing or enriching.
_____ 5. Very enriched. Along with appropriate responses to other speeches and initiation of required speeches, the child displays adeptness at turntaking and topic continuity and makes many speeches that are not necessary, but socially relevant.

e. Does the child seem to use logic and associative principles or laws in understanding or explaining various things?

_____ 1. No; no instances were observed.
_____ 2. Yes; one or two instances were observed.
_____ 3. Yes; three or more instances were observed.
_____ 4. Yes; child often seemed to invoke a "logical" or "associative" or "rule-governing" principle.
_____ 5. Yes; the child very often employs a logical or associative principle and is quite adamant concerning these.

Figure 2 (cont'd)

f. Child <u>actively describes</u> things, people, and ideas during the Head Start day.

Scale:	Dyads:
1 Rarely or never	_____Teacher-child
2 Sometimes	_____Aide-child
3 Fairly often	_____Other children-child (same sex)
4 Often	_____Other children-child (opposite sex)
5 Very often	_____Self (child)
	_____Overall (child)

<u>Notes for Figure 2</u>

1. Filled out for each target child each week of classroom-based field work.
2. Used for teachers (in revised format) as well.
3. Repeated for Spanish and Mixed E-S.
4. Contexts (story time; free play; etc.) can be substituted for dyads.

ratings will provide important evidence of change and of the effectiveness of each model in producing language change over time. Ratings can also specify which context or dyad was most likely to lead to increased sophistication of the child's utterances, greater talkativeness, etc.

(It should again be emphasized that Figure 2 is not a complete list of the ratings which will be developed from the models' own specifications, and from the contract. It is simply an example for purposes of illustrating the participant researcher's data collection techniques in the field.)

Figure 2 uses dyads to illustrate one kind of context-specificity in language use. Dyads could be replaced with contexts themselves to aggregate the ratings. For example, the language use of a particular teacher in (a) free play, (b) reading time, or (c) object-manipulation situations could be rated. Ratings can also be used to judge teacher effectiveness. The point is that Figure 2 illustrates a <u>general</u> strategy for using informed participant researcher ratings.

The event sample methods of direct observation and the participant researcher weekly ratings are designed to complement each other. The ratings will usually be at a higher level of generality and will require more global overall assessment of a child or teacher or setting than will the specific event sample. Although it is possible to aggregate the event sample data into overall ratings, this is not always appropriate. Furthermore, the event sample ratings will be done at different times of the day or week on different children, whereas a great deal of additional knowledge of children's language use will be available to the PR at other points of the day which can be used in forming the rating judgment at the end of the week. This is not possible using the

straight event sample format.

Still another rating task for the PR will be a simple rank order procedure. An example of this would be to take all of the target children and teachers in the class and make a judgment ranging from most to least talkative, or a determination of the children with the highest judged English or Spanish language skills down to those with the least judged English or Spanish language skills. We will ask the PR to make such assessments on a rating scale at the end of each week--thus summarizing an overall assessment of the domain of language use for that week. These ratings and assessments may change over the year, and their degree of change or stability will be an important consideration.

(3) Ethnographic notes. In addition to the focused event sample style observations, and the ratings of teacher and children, the PRs will be keeping a series of notes on topics relevant to the overall functioning of the school and the implementation of each model. A series of topical domains or an index and mnemonic system will be developed for this purpose. Some examples of topics which will be included in these notes are teacher morale; advisory committee functioning; role of the parents in the school; participation of aides and volunteers in the classroom; weekly attendance notes and records on attendance for each child; comments by the PR on his/her role and influences if any on the teachers in the classroom situation; and any exogenous difficulties the school is having with the community (i.e., problems with staffing, buildings, politics, funding, and so forth). In addition, a series of general domain categories will be developed, related to the central goals of each model (i.e., biculturalism, language switching, cognitive style differences,

etc.). These notes will be typed and submitted to Juárez and Associates at the end of the two-week in-class observational period.

These focused field notes, specific to domains of interest to the overall evaluation and to each model, should provide an important qualitative assessment of the functioning of each school, its implementation of the model, and any exogenous factors which are likely to have influenced either the in-class observations and ratings, the pre- and post-test measures, or the experimental arrangement.

(4) Phone-call updates. The PR will make phone calls periodically during the year to maintain contact with the family sample identified at the beginning of the year at each Head Start site. The PR will fill out a brief phone call checklist of information through the use of telephone updates to monitor any changes in residence, child caretaking, educational experiences for the child (other than Head Start), census information, etc. This is an inexpensive method of obtaining basic information on the control and school family samples.

(5) Teacher interviews. The participant researcher is also to collect information by means of a set of questions regarding the attitudes, knowledge, and behaviors of those teachers and teacher aides employed at the Head Start centers. The information called for is to be secured by the participant researcher through conversations with and observations of teachers and aides during their normal teaching duties. Because of the interest in learning about changes in teachers' attitudes, knowledge, and behaviors during the program's implementation, the participant researcher is instructed to collect information related to the questions as early and as late in the program as possible.

(6) Teacher ratings. Although the participant researcher will be making his/her own ratings of language use and other variables, s/he may also take advantage of the knowledge of the Head Start teachers and aides. We may present a simplified rating task and ask the teachers to complete this at various points during the year. These teacher ratings of language skill or of other aspects of the program (e.g., implementation), made as specific as possible and as parallel as possible to those of the PR, will be extremely useful as independent expert judgments of children's language skills and their change over time.

(7) Psychometric testing. The participant researcher will be responsible for the coordination of the pre- and post-testing at his or her assigned site. The specific procedures to be followed for both testing and the recruitment and training of a team of local testers at each site are currently being developed and piloted. A document describing all of the procedures to be followed will be circulated by Juárez and Associates to the curriculum model developers for formative comments and suggestions.

III. Data Management

Juárez and Associates plans a series of data collection points for assembling and monitoring the weekly notes, classroom observations, home observations, interviews, etc. Once a month each PR will submit to Juárez and Associates all the materials collected to that point (see Appendix A). These will be reviewed by the supervisor in Los Angeles, and phone calls and other assessments and monitoring information will be fed back immediately to the PRs to make sure that they are completing the materials correctly. We consider this periodic monitoring of field data essential to the comparability, reliability, and completeness of the data. We will also use this technique as a way to monitor the performance of the PR long distance. We feel that the most effective way to monitor accuracy, time specificity, and amount of data is to require the information to be submitted regularly by each field site.

Monitoring Data Quality

It is important to calibrate, validate, and obtain reliability on the data collected by the PRs in observations at the four different sites. We plan to approach this task in a number of ways, depending on the type of data, focus of interest in the topic, and cost. Interview and survey materials will be standardized. Training meetings to ensure common definitions of terms and common interview procedures will be utilized (see Appendix A).

Classroom observations will be calibrated in several ways. Specific class settings will be selected for focused observations -- examples include teacher-child dyads; small groups of children; small groups of children and a teacher; children in free play outdoors; etc. A training session will be conducted prior to classroom observations to provide baseline reliability for standardized observations. The PR supervisor will make field survey visits to handle definitional and measurement problems and to do parallel observations for reliability checks. Frequent phone call checks will be made with PRs to monitor progress in observations. Clearly defined baseline items will be included in each PR's observation schedules. These items will be carefully defined and related to specific contexts, and will be used by all field observers and by the PR supervisor in parallel observations.

The goal of a PR is to provide reliable enough data to complement the high validity of data s/he is expected to provide. The relevant reliability comparison to the classroom observational data PRs will collect is not formal tests or one-shot field visit observations; the context-specific, long-term, quantitative/qualitative mix of data the PR will collect trades validity and multiple measures for the formal reliability of tests. We believe acceptable reliability is attainable in these measures, but that constant vigilance and planning are required.

Reactivity of participant researcher data. Related to data quality, reliability, and validity issues is the problem of reactivity. Each of the PRs is going to have individual, unique

influences on schools where s/he will be working, the families s/he will be contacting, and the models and model developers with whom s/he will be working. The fact that there will be such reactivity effects and that the individuals chosen as PRs will influence and be influenced by the setting they are in is undeniable. This leads to two important points about reactivity. First, it is clearly a trade-off. Our evaluation strategy trades (ostensibly) impersonal and less reactive periodic outsider visits to field sites for purposes of testing and interviewing, for greater reactivity from the PRs coupled with obviously greater inside knowledge and information about the children, teachers, and parents.

It should be noted, however, that outside evaluation visits and the qualities and characteristics of the outside evaluators (their sex, their age, their response to the setting, the amount of time they are there, and so forth) are also reactive, and that these factors also have an unknown differential and uncontrolled effect on the situations in the various field sites. Indeed the very design of such a periodic shotgun evaluation visit technique has a very reactive effect -- the intensive preparation for evaluation visits, the unknown influence of the visit itself, followed by a (variable) period of no visits and no contacts. This is a well-known problem in evaluations of this kind. The reactivity of a PR in a site is therefore only a <u>relative</u> issue: there is reactivity in all techniques; the question is, is it worth trading the kind of long-term, habitual presence of a PR and the data that s/he will gather for the <u>different</u> kind of reactivity that s/he will produce? Juárez and Associates believes that the trade

is worthwhile, since both the contract and staff knowledge of the kind of evaluation needed in the situation indicate that a contextual, quantitative/qualitative mix using a PR is the strategy of choice.

The second issue in reactivity involves being able to measure and control it. There are ways to compare the reactivity of different PRs by having them systematically include such information in their own notes and by having teachers and other staff make judgments about their effects. Obviously this is only a partial solution, although the use of several weeks of systematic observation and test procedures as validation for PR observation will assist in judging the problem of reactivity. We believe that we can measure to some degree the amount of reactivity that will be occurring. In addition, some data are relatively low reactive in all situations and some are relatively high reactive. We can test whether low reactive measures are stable regardless of the person and of the setting where the measures are being collected.

IV. Data Analysis: Uses of Qualitative Notes

Domain and Indexing System

As is indicated in the data collection sections and in the chart depicting the participant researcher's work schedule over the year (see Appendix A), an important part of the participant researcher's time will be taken up with the writing and submission of field notes to Juárez and Associates. The field notes will be written according to a series of topical headings which will provide a set of domains related to the contract-specified variables of interest, including variables specific to each model. The field notes will be structured and indexed as they are being written. The indexing system will permit frequent review and comparison of the field notes by the field supervisor at Juárez and Associates and by the participant researcher while in the field. The system of indexing and domain categories will be directly tied to the topics which will be covered in the preparation of final pamphlets describing each model and its implementation at each site.

There will also be an opportunity for participant researchers to make notes and observations not covered in the index and domain system, since it is certain that many topics of interest will be unique to a given site, or will emerge after the participant researchers are already in the field. However, these will be a supplement relevant to the preparation of each pamphlet and for interpretation of results for each site; the core of the notes will consist of comparably collected materials submitted on a weekly basis while in the field.

Mid-Year and Final Reviews of Field Notes

Halfway through the year all the participant researchers will assemble in Los Angeles for observational reliability purposes and for comparison and review of their field notes. At this time each PR will write brief summaries of his/her overall assessments up to that point. This review will help in focusing and providing a synthesis and summary of the notes which have been collected. At the end of the year when the participant researchers again assemble at Juárez and Associates' Los Angeles offices, these reviews and summaries will again be done as part of the final reports required of the participant researchers. This will provide Juárez and Associates with a clear set of summarized qualitative notes assessing the key areas of implementation, classroom impact, and learning, and outlining the control group tracking which will have been undertaken throughout the entire year. A further revision and collapsing of the domain and index categories will be undertaken at this time. It is anticipated that a set of core concerns relevant to the contract and to the objectives of Juárez and Associates' evaluation will be completed across all sites, whereas certain unique domains with more limited relevance will be worked on based on the participant researchers' assessments of the situation at the midpoint and conclusion of the evaluation year.

Interpretation of Quantitative Results

One of the important uses of the field notes data will be to assist in interpreting the quantitative pre- and post-test results and the participant researcher ratings and context-specific data collection procedures planned at each site. The use of field notes for interpretation and understanding in a systematic way is vital to the goal of the contract; this goal is not simply to ask if changes occurred, but to ask how the changes occurred and the reasons for these changes based specifically on each model and its strengths.

Some hypothetical examples might illustrate some of the uses of qualitative field notes for this interpretative and understanding function. Suppose that at a site measures of English language production show no pre- and post-test differences. Assume that this model had as one of its goals the encouragement of bilingual language use in the classroom. Why was there no difference between pre- and post-test results? An examination of the implementation categories relevant to the encouragement of English language use is the obvious first place to look. Perhaps the notes indicate that this model goal was really not implemented effectively, and the field data and teacher interviews confirm this. The interpretation would be that the treatment effect did not actually occur in the classroom; therefore, the comparison of the (supposed) treatment to a control is invalid in this case.

Field notes may however show more interesting patterns. They may indicate, for instance, that only certain teachers at

this site utilized English while others did not. The notes may indicate some of the reasons for this: e.g., personnel choices and changes, different cultural backgrounds of staff or children, and so forth. These field note observations may in turn lead to a look at the participant researcher ratings or classroom observations of those teachers who did apparently implement the model goal in the area of English language usage compared to those teachers or situations where it was not implemented. If it were ultimately discovered that those particular teachers or situations with effective implementation did in fact lead to a difference between the treatment and control groups of children, a clear implication of effectiveness of this model would be strengthened and both the interpretation of the statistical results and the write-up of the pamphlet describing implementation for this model would be influenced as a result.

Field notes may also indicate variability in the program over the year. A comparison of these assessments of variability and diversity in the school program to the periods when pre- and post-testing was done may discover unusual circumstances in the timing of testing which influenced the test results. Field notes on pupil and teacher absenteeism and attendance may have a profound effect on the pre- and post-test results as well. A comparison between the particular days or the number of days when children and teachers were there in the classroom and the scores on pre- and post-tests of language usage or social-emotional and concept development may reveal that those children and teachers who were present a great deal, and who were present during times

when the model was being effectively implemented (as evidenced by the field notes and classroom observations) may have changed compared to the control group. This specification of samples for use in final data analysis would go beyond a comparison of all of the treated children to all of the control children, or all of the Head Start parents to all of the non Head Start parents.

Exogenous Factors

It is a well-known phenomenon that factors exogenous to the school may prove to have the most effect upon the children, the teachers, and the Head Start institutions themselves. Political changes in the community, funding uncertainties, health or related problems, attendance fluctuations, relationships with outside school systems or other preschools in the area -- these are all relevant in interpreting the overall comparisons between the control and Head Start children. Examination of such factors will only be possible through a continuous monitoring of domain and index categories relevant to these topics noted by the participant researchers throughout the year. These can then be used by the PRs and project directors in assembling final reports and interpretations of the effectiveness of each model at its two sites.

Generation of New Hypotheses

Another important use of field notes is to discover new ideas and relationships to be tested other than those originally thought important. Participant researchers at several different sites, for example, may all observe in their field notes that

schools with a high degree of participation on the part of parents as active teacher aides (rather than as members of an advisory committee) appear to have high morale, better attendance, and increased student and teacher involvement. This is turn may lead to an examination of sites with or without this feature.

Several participant researchers may note that schools that have male as well as female teachers and staff appear to have less aggression and more attentiveness on the part of the children and more structure in the programs. These observations may in turn lead to data analysis comparing male and female teachers and looking at overall program consistency and student attentiveness according to these variables. The presence of male teachers may in turn interact with model goals which in turn may have been somewhat altered in actual implementation according to the sex of the teachers.

Conclusion

These are purely hypothetical examples to illustrate some of the ways in which the field notes can be used -- for interpretation of quantitative results; for a better understanding of the pattern of results across sites; as crucial data for the assessment of implementation; for the generation of new hypotheses for analysis; and as suggestions for re-analysis of results that do or do not find differences between treatment and controls.

Integration of Psychometric and Ethnographic Data in Statistical Analyses

V. Analysis of Qualitative and Quantitative Data

Certain data from the ethnographic observations are in the form of ratings, time estimates and counts. These data can be integrated within the general analysis plan for the psychometric data (tests).[1] The combined data will be analyzed separately for each curriculum model; thus, the data analysis for the general case can be discussed as if there were one model. Analysis for these ethnographic data will proceed at two levels. At the first level, ethnographic observations will be combined with psychometric data as a multi-index of an underlying trait such as language comprehension and production in the child's first and second languages. Thus, observational data in conjunction with psychometric data will be used to acquire a better estimate of the effects of the treatment from pre- to post-test. The second level of analysis involves analyzing the ethnographic observations separately over four points in time. We can examine observations of, for example, children and teachers to estimate change within and between replication sites.

In addition the following issues will be addressed by the use of ethnographic observations:

1. Verification of language preference categories identified by psychometric instruments.
2. Treatment of ethnographic observations in Spring as separate dependent variables and observations in Fall as covariates to acquire a better estimate of the effects of the treatments.

---

[1] See Juarez and Associates' Design and Analysis, October 1978.

3. Unit of analysis: ethnographic observation can be used to identify alternative units of analysis. For example, many of the classrooms will have children from different language groups (i.e., Spanish dominant and English dominant). Thus, the social groups and type of interaction among children within the classroom may differ depending on each child's language preference. Thus, one definition of the unit of analysis might be the social group rather than the individual student or the class itself.
4. Ethnographic notes can be used in two ways in the analysis: (1) to interpret the data and (2) to help explain the data by generating hypotheses about the relationship between the dependent variables (outcomes) and independent variables (classroom activities).

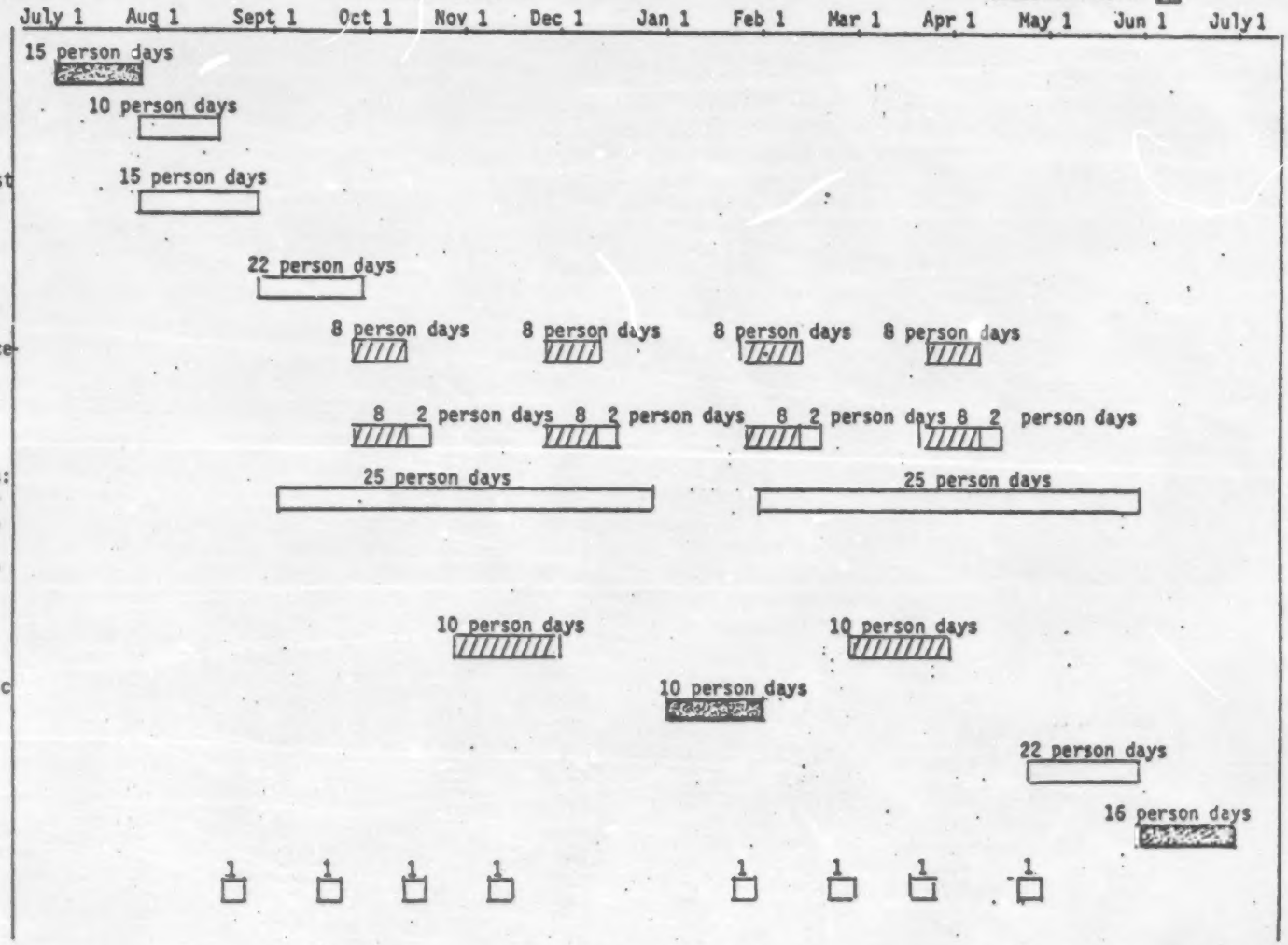
Appendix A

An Evaluation of Bilingual Bicultural Head Start Curriculum Development Models

Ethnographic Component's Participant Researcher: Task Analysis

1979-1980

In Los Angeles
Field Work
Classroom Work



Appendix B

Cross-Model Behaviors

Domain: Spanish/English Language Production

Component: Oral Language Development-Descriptive Skills

a) ability to identify common objects

(Examples: Classroom items such as play dough, blocks, crayons, paints, etc. Food, clothing, transportation vehicles, objects in the community (i.e., signals, street lights, houses, apartments, etc.), plants, animals, furniture, etc.)

b) ability to describe common objects and situations

(Examples: Classroom events (i.e., roll call, lunchtime, naptime, grouptime, storytime, pick-up time, etc.) Daily routines (breakfast, dinner, etc.) special holidays)

c) ability to describe self

(Examples: Body parts, physical dimensions (i.e., tall/short; fat/thin)

d) ability to describe others

(Examples: Body parts, physical dimensions (i.e., tall/short; fat/thin)

e) ability to express feelings

(Examples: Hunger/satiation, happy/sad, hurts/doesn't hurt, sleepy/not sleepy, feels good/doesn't feel good, anger, jealousy)

Component: Oral Language Development - Conversational Development

a) ability to give verbal instructions

(Examples: Directing another on how to perform a specified task, i.e., climbing jungle gym, telling someone where an object is, i.e., under the table)

b) ability to make up(relate) a story or event; alternative - can make statements about favorite story or event

(Examples: Tells another about a special occurrence, i.e., birthday, special outing, T.V. program)

Component: Oral Language Development - Linguistic Markers

a) ability to use the negative forms in statements

(Examples: Don't, can't, won't, have not, etc.)

b) ability to use interrogative forms by intonation

(Examples: Are you going? Me?)

c) ability to pluralize regular nouns

(Examples: Cup/cups, book/books, shoe/shoes)

d) ability to produce simple sentences

(Examples: I want some juice.)

e) ability to produce compound sentences

(Examples: I want some juice and I want some cookies too.)

f) ability to produce complex sentences

(Examples: I want some juice but there isn't any more).

g) ability to use present tenses

(Examples: El conejo brinca.)

h) ability to use past tenses

(Examples: El conejo brincó.)

Domain: Spanish/English Language Comprehension

Component: Word Meanings

a) identifies spoken words with correspon ing pictures

Component: Auditory Discrimination

a) identifies common sounds from the home, school and community

(Examples: Animal noises, street noises)

b) identifies familiar voices

c) distinguishes non-vocal sounds which are alike in a given way

(Examples: Rhythm, intensity, pitch, noises)

d) identifies rhyming words

(Examples: True/blue, tool/pool, cambiar/llevar/sentar, robo/lobo)

Component: Determining the Main Idea

a) recall outstanding events from an oral story

(Examples: T.V. program)

b) able to recall what s/he did in the classroom/home

c) able to provide details about classroom/home

Domain: Concept Development

Component: Visual Discrimination

a) identifies likenesses and differences (color, size, shape, and texture) in objects

Component: Seriation/Sequencing

a) arranges objects in sequence

(Examples: Small to large, short to tall, thick to thin)

b) describes the relationship of sequenced items

Component: Classification/Grouping Objects

a) sorts and matches objects (size, form, condition, shape, color, substance)

b) holding more than one attribute in mind at the same time

(Example: Can you find something that is red and made of wood?)

c) distinguishes "some" "all" "more" less" and "none"

Component: Spatial/Relationship/Space

a) children use terms showing a relationship to time

(Examples: "Mas tarde" "Hasta luego" Times of the day (morning, afternoon, evening) after, before, sometimes, etc.)

b) understands "time" unit or ideas

(Examples: Uses time terminology in a way that demonstrates understanding.)

c) observes that clocks and calendars are used to mark the passage of time

(Examples: In the housekeeping center, children note that clocks, watches, and calendars are used to designate time.)

Component: Symbolic Representation

a) uses materials symbolically

(Examples: Uses blocks to make an apartment, house or to represent people, animals or objects. Use playdough to make "make believe" foods, animals, objects, etc. Uses materials in the community (rocks, leaves, twigs, etc.) to represent items such as cars, people, houses, etc.)

b) makes drawings and drawings

c) imitates actions, sounds

(Examples: Imitates the sound of a fire alarm, imitates the movement of a favorite T.V. personality (The Fonz), imitates movements made by others (i.e., bending, twisting, jumping))

d) pretends

(Examples: Child can assume the role of someone else, i.e., child can take on the role of people, animals (make believe (pantomime) you are a bunny rabbit))

Domain: Socioemotional Development

Component: Socioemotional Development

a) sustains interest in group activity

(Examples: Child completes a group game such as "Ring Around the Rosy." Child shows interest in group activity for a determined period of time (given contextual restraints))

b) works cooperatively with others; alternative - child cares for her own needs without continual adult assistance

(Examples: Child appears to enjoy playing with others in a cooperative manner. Given contextual restraints, child gets adult attention in ways that are acceptable for age.)

c) follows directions

(Examples: Child makes an effort to follow the social rules (i.e., keeping hands to self, waiting for turn).)

d) expresses desire to attend the Head Start Program

(Example: spontaneous pronouncement)

e) shares and takes turns

(Examples: Child is willing to take turns when the need occurs, allows others to share her materials.)

f) evidences pride in accomplishments

(Examples: Indicates happiness or feelings of self-worth over completed school task.)

g) sees self as capable of mastering new situations

(Examples: Expresses intent to confront new situation/tasks "I can do it.")

h) communicates capabilities to others

(Examples: Reports confidence concerning home/school task.)

JUAREZ & ASSOCIATES
DOMAIN - SPANISH/ENGLISH LANGUAGE PRODUCTION
DRAFT

COMPONENT - ORAL LANGUAGE DEVELOPMENT-
DESCRIPTIVE SKILLS

BEHAVIOR	Setting/Context: H.S. CENTER				Setting/Context: Playground			Method: H.S. CENTER				Method: HOME		
	Adult-small group	Adult-one to one	Peer group	Other, Specify*	Adult-child	Child-sibling or friend	Other, Specify*	Teacher obs.	P.R. obs.	Teacher Interview	Teacher's Records	Parent obs.	P.R. obs.	Parent Interview
a) Ability to identify common objects (e.g., food, classroom, items, playground items, clothing, transportation vehicles, items in the neighborhood; signals, streetlight, houses, building, furniture).														
b) Ability to describe common events and situations (e.g., times of day, special holidays, classroom events, such as roll call, lunchtime, naptime, grouptime, storytime; pick-up time).														
c) Ability to describe self (e.g., body parts, physical dimensions; tall/short, fat/thin).														
d) Ability to describe others (e.g., body parts, physical dimensions; tall/short, fat/thin).														
e) Ability to express feelings verbally (e.g., hunger/satiated, happy/sad, hurts/doesn't hurt, feels good/doesn't feel good, sleepy/not sleepy, angry/not angry, jealous/not jealous).														

Appendix D

Participant Researcher

The ethnographic plan requires that an on-site participant researcher (PR) be placed in each of the eight demonstration Head Start Centers. The participant researcher should be familiar with both field/naturalistic and controlled research settings. The participant researcher's primary role definition and responsibilities involve collecting of both qualitative and quantitative evaluation data. Data collection procedures include formal and informal observation, interview techniques, supervision and administration of psychometric tests, and compilation of field notes and daily logs. In addition, the participant researcher will be responsible for data reduction, writing reports, and coordination of on-site evaluation activities.

Because of the data collection responsibilities of the PR, training in the social sciences is required and research experience in education is strongly preferred.

I. EDUCATIONAL BACKGROUND PREFERRED

A minimum of an MA degree in the social sciences (i.e., anthropology, education, linguistics, psychology, or sociology), with an interest in child development, early childhood education, and bilingual education. Preference will be given to doctoral students in the above mentioned academic disciplines.

II. EXPERIENCE PREFERRED

A. Since the PR will be overseeing a bilingual/bicultural program in which the majority of children will be Spanish speakers, it is necessary that the PR be bilingual, biliterate, and bicultural.

B. Appropriate relevant experience in child development settings, particularly preschool contexts.

1. experience with young children, ages 3-5
2. experience with teachers and parents of young children and/or appropriate personnel who assume responsibility for young children.

C. Experience with field or experimental research focusing on young children.

D. Experience with community-based social-service delivery programs.

III. RESPONSIBILITIES

A. Maintain field notes according to a semi-structured set of domains, and a running account of the day-to-day activities in the Head Start Center.

B. Interview teachers and administrators at the site where they are working.

C. Administer or supervise the administration of formal psychometric pre- and post-test measures.

D. Conduct formal and informal observations in each Head Start Center of selected children and selected contexts.

E. Conduct home observations and parent interviews in non-Head Start comparison group of families and sub-sample of HS childrens' home.

F. Interview comparison-group families, including phone-call updates.

G. Act as a liaison between evaluator (Juárez and Associates) and Head Start site.

H. Assist in initial recruitment of children at each Head Start site.

IV. SALARY RANGE

$12,000-$15,000 per year plus medical and dental insurance

Appendix E

Coordinator of Participant Researchers

It is essential that a full-time coordinator be available to coordinate all of the participant researcher (PR) activities, data collection, data handling, data analysis, reliability analysis, and communication with model developers. Field coordination and a central source to ensure comparability of field data is critical.

I. EDUCATIONAL BACKGROUND

The coordinator must have a Ph.D. in either anthropology, education, linguistics, psychology, or sociology. S/he must have had extensive experience with both qualitative and quantitative data analysis techniques, and experience in educational settings is highly preferred.

II. EXPERIENCE

A. Appropriate experience in child development settings, particularly preschool contexts

1. Experience with young children, ages 3-5
2. Experience with teachers and parents of young children and/or appropriate personnel responsible for young children.

B. Experience with community based social-service delivery programs

C. Experience with the development and conduct of field work in educational settings.

D. Experience with the development of procedures for informal and formal quantitative and qualitative observational methods.

III. RESPONSIBILITIES

A. Participate in the selection and assignment of PRs to each Head Start site.

B. Participate in the initial and ongoing training of PRs

C. Supervise the PR's recruitment of children at each Head Start site.

D. Schedule and coordinate PR activities at each Head Start site.

E. Monitor reliability of field observations and other data collection.

F. Be responsible for the assessment of implementation data of each model in the field site.

G. Be responsible for writing appropriate progress and final reports.

IV. REQUIREMENTS

Since the coordinator is responsible for overseeing data collection of preschool bilingual, bicultural programs, it is necessary that s/he be bilingual/bicultural in Spanish and English. S/he must also be willing to travel.

V. SALARY

$18,000-$20,000 per year plus medical and dental insurance.

Bibliography

Brown, A., and Campione, J.C. Permissible inference from outcome of training studies in cognitive development research. The Quarterly Newsletter of the Institute for Comparative Human Development, 1978, 2 (3).

Campbell, D.T. Qualitative knowing in action research. Kurt Lewin Award Address, Society for Social Issues, September 1974.

Cole, M., and Scribner, S. Culture and thought. New York: John Wiley, 1974.

Cole, M., Gay, J., Glick, J.A., and Sharp, D.W. The cultural context of learning and thinking. New York: Basic Books, 1971.

Cole, M., Sharp, D., and Lave, C. The cognitive consequences of education: Some empirical evidence and theoretical misgivings. The Urban Review, 1976, 9, 218-233.

Cummins, J. Psycholinguistic evidence. In Center for Applied Linguistics, Bilingual Education: Current Perspectives, November 1977, pp. 78-89.

Erickson, F. Some approaches to inquiry in school-community ethnography. Anthropology and Education Quarterly, 1977, 7 (2), 58-69.

Erickson, F., and Schultz, J. When is a context? The Quarterly Newsletter of the Institute for Comparative Human Development, 1977, 1 (2), 5-10.

Juárez and Associates, Inc. Evaluation of bilingual/bicultural Head Start curriculum development project. The initial phase: A cooperative effort by educators and evaluators, Proceedings of the San Antonio Bilingual Head Start Conference, San Antonio, Texas, November 15, 1977.

Juárez and Associates, Inc. Review and recommendation for the test battery. An evaluation of the Head Start bilingual bicultural curriculum development project. Report prepared for the Administration for Children, Youth and Families, Contract No. HEW-105-77-1048, July 1978.

Juárez and Associates, Inc. Revised evaluation design and field plan. An evaluation of the Head Start bilingual bicultural curriculum development project. Report prepared for the Administration for Children, Youth and Families, Contract No. HEW-105-77-1048, April 1978.

Juárez and Associates, Inc.

AN EVALUATION OF THE HEAD START BILINGUAL BICULTURAL CURRICULUM DEVELOPMENT PROJECT

Addendum to: A Qualitative/Quantitative Data Gathering Approach

Authors: Thomas LaBelle, Luis C. Moll, Ray S. Pérez, Thomas Weisner, Marlene Zepeda

February 1, 1979

Contract No. HEW 105-77-1048

Project Director: Luis C. Moll, Ph.D.
Co-Project Director: Ray S. Pérez, Ph.D.

Prepared for: Dr. Steve Martinez,
Project Officer
Administration for Children, Youth, and Families
Office of Human Development Services
Department of Health, Education, and Welfare

PS011560

1100 Glendon Avenue • Los Angeles, California 90024 • (213) 478-0826 • (213) 879-5899

Acknowledgments

A plan of this scope and complexity requires the creativity and technical expertise of many individuals. We would like to acknowledge their contributions. Drs. Martha Bernal, Susan Loucks, Barry McLaughlin, Ronald Gallimore, Hugh Mehan, and Jane Stallings. In addition, we would like to thank Drs. Richard J. Shavelson and Thomas Weisner, who have continued to provide us with their technical guidance. We also thank Ms. Sandra Green and Ms. Dee Meyers, who helped prepare the report, and Ms. Margaret Mason, who gave editorial guidance.

TABLE OF CONTENTS

I. Introduction 1

II. Review of Evaluation Objectives 3

III. Design and Methodology 4

IV. Scope of Work 7

A. The Participant Researcher 7

B. Data Collection Activities 9

1. Assessment of Implementation 9

2. Assessment of Program Impact 13

a. Assessment of Child Impact 14

b. Assessment of Parental Impact 18

c. Assessment of Teacher Impact 21

V. Data Management 23

VI. Data Analysis 24

LIST OF APPENDIXES

Appendix A - Time Line of Activities for the Evaluation

Appendix B - Summary of Data Collection Activities

I. Introduction

This proposal is submitted to fulfill the requirements under Amendment Number 1 of HEW Contract 105-77-1048. The purpose of this modification is to prepare a plan for gathering ethnographic (qualitative) data in the Evaluation of the Bilingual/Bicultural Head Start Curriculum Development Project. Rather than attempt to rewrite the original reports submitted by Juárez and Associates to the Administration for Children, Youth, and Families (ACYF; see Juárez and Associates, July 1978 and December 1978a), it was decided to prepare the enclosed addendum or scope of work to accompany these reports.

This addendum seeks to accomplish the following goals: (1) to summarize and integrate both the qualitative and quantitative aspects of the evaluation design and describe all the procedures recommended for each remaining phase of the contract and (2) to provide revisions to the original reports based upon comprehensive discussions with members of the ACYF staff, the curriculum model developers, members of the evaluation's advisory committee, and Juárez and Associates' staff, and on the nearly 100 pages of recent comments submitted by the nine consultants on the qualitative aspects of the evaluation design.

This document is divided into several sections: a brief review of the evaluation objectives, followed by discussions of the design and methodology, the participant researcher role, the assessment of implementation, and the assessment of program impact. In some instances, changes proposed here (e.g., a reduction

in the number of field workers) supersede information contained in the original documents. Therefore, each of the principal sections contains a description of the tasks we propose to carry out during the evaluation. The reader should note that the order of task presentation is not necessarily the order in which we plan to carry out the tasks. For this reason, all of the proposed tasks are outlined in chronological order in Appendix A.

II. Review of Evaluation Objectives

The work statement in the evaluation contract with Juárez and Associates specifies that the purpose of the evaluation is to assess the effectiveness of four early childhood bilingual bicultural curriculum models for Spanish speaking children. More specifically, the contract requires the evaluators to collect information on the following:

(1) The feasibility of successfully implementing the models in more than one setting. The evaluators must secure information that will assist others in learning about the potential of the model for implementation elsewhere.

(2) The extent to which the models, once implemented, are meeting their objectives. This means that one emphasis in this evaluation is on measuring the change in children, teachers, and parents as a result of their participation in the four curriculum models.

(3) The extent to which the models are greeted favorably by Head Start staff and parents. This means collecting information regarding staff satisfaction with their participation in implementing each model and parents' attitudes and opinions regarding the curriculum model.

Finally, a set of pamphlets will be developed that summarize all the information that is to be collected above. The purpose of the pamphlets is to aid in the dissemination of information about the four models to interested Head Start programs.

III. Design and Methodology

The evaluation contract specifies the use of an experimental pre- and post-test design. Children are to be assigned randomly to treatment and control (no-treatment) groups. Subsequently, at the beginning and conclusion of the treatment, the children will be tested on selected competency measures (see Juárez and Associates, July 1978). Parents will be administered questionnaires which assess background characteristics, attitudes, and knowledge with regard to education in general and bilingual education in particular, and attitudes toward educational and vocational careers for their children. Head Start teachers will also complete a questionnaire designed to assess their background, their understanding of the terms "bilingual" and "bicultural" in the context of an early childhood program, their attitudes toward children, sensitivity to children's special characteristics, and so on.

The evaluation approach developed by Juárez and Associates integrates both qualitative and quantitative methodologies as complementary data-collection strategies while retaining a basic experimental design. That is, the mix of both a qualitative and quantitative evaluation within the framework of an experimental design is seen as feasible and desirable.

In order to facilitate the integration of both methodologies, a time-series approach to data collection has been adopted. This arrangement permits the administration of pre- and post-tests as well as periodic and systematic ongoing classroom observations; it also allows the field worker sufficient time for both analysis

and writing, as well as performing the actual field work. Basically such an approach involves the testing of program participants at the beginning and conclusion of the treatment, plus classroom observation for four specified and preselected two-week periods. (See Juárez and Associates, December 1978a, Appendix A, for approximate dates.) These naturalistic classroom observations focus on samples of approximately 15 children per site. The sample can be stratified, for example, by the children's language preference, sex, and/or age. Limiting the observations to a sample of children instead of observing every participant makes a difficult task manageable and facilitates the use of multiple, but complementary, data-gathering methods.

In our view, in order to carry out a multi-method, contextual evaluation study that combines both quantitative and qualitative methodologies, it is essential to carefully train and place at the evaluation sites a field observer or "participant researcher" (see Juárez and Associates, December 1978a). These on-site evaluation staff members will oversee the psychometric data-gathering and will become the primary people responsible for collecting data on both the school and the control group children.

Juárez and Associates has carefully reviewed the comments of the consultants, advisory committee members, and ACYF staff who, in general, felt that the management and reliability problems inherent in hiring, training, and supervising eight participant researchers (one full-time field worker at each of the eight replication sites) would be formidable. We are also aware that

in just five months, these individuals would have to be recruited, trained, and available for placement. Consequently, Juárez and Associates now proposes that a participant researcher be placed full-time at one replication site of each curriculum model. That is, four participant researchers will be utilized for the implementation and contextual data collection effort, and for the monitoring of the control sample of parents and children. At the remaining four sites, trained and experienced observers (Juárez and Associates staff members) will make quarterly visits in order to assess implementation and the general functioning of the program.

By reducing the number of participant researchers from eight to four, and by having these field workers concentrate on one site per model, field management can be more effectively controlled. This will permit increased reliability checks, closer monitoring of data quality, and frequent field and office meetings between the supervisor and the participant researchers. The overall results of the study should be strengthened by concentrating on four in-field, full-time site assessments and four implementation-only site assessments.

IV. Scope of Work

A. The Participant Researcher

There are several reasons why Juárez and Associates advocates a participant researcher (PR) role. It is our position that too many evaluations utilize outsiders who come into the classroom to administer tests or questionnaires or to interview and then leave. A common complaint is that this type of evaluator does not ask questions relevant to the programs or the sites and fails to obtain a fair sample of the typical life within the school. Such an evaluator cannot really learn in the short time available what is going on and cannot know nearly as much about the system as the insiders who run the school.

A PR is both an insider and an outsider; his/her role helps to overcome this structural problem of differentiating between the roles of "outsider" evaluator and "insider" staff member. His/her participation can provide revealing contextual and relevant data at each site. Thus, the PRs have an important and difficult role. For selected periods of time they are simultaneously involved in the day-to-day activities of each Head Start center and gather inside data (on teacher activities in the classroom, how parents and children react on a day-to-day basis, etc.). But PRs are also responsible for the coordination or supervision of the collection of formal test measures, for providing controlled feedback to the field sites, for interviewing the parents and teachers, and for contacting a sample of homes to gather ongoing information on educationally relevant activities.

(For further discussion of the participant researcher strategy, see Juárez and Associates, December 1978a.)

The following is a list of tasks that must be completed in order to implement the proposed PR strategy:

(1) Recruitment, Training, and Placement of Participant Researchers. Juárez and Associates will advertise for, hire, and train four bilingual participant researchers to perform intensive ethnographic field research and data collection at four field sites. The four field workers will be placed at their sites in early Fall 1979. These researchers will be trained in direct observation of naturalistic behavior in the classroom, field work techniques, role management in the field, and ethnographic note taking procedures. (See Juárez and Associates, December 1978a, for the job description and the training plan.)

(2) Management of Participant Researchers. Juárez and Associates will hire an experienced full-time supervisor who will coordinate all of the qualitative data collection and thereafter the write-up and analysis of these data. This individual will be responsible for implementing the participant researcher field plan, monitoring data quality and reliability, and writing up the results. (See Juárez and Associates, December 1978a, for a description of the coordinator position.)

(3) Quality Control and Reliability. Juárez and Associates will develop and implement a plan for assuring the adequate reliability of observational data and the completeness of ethnographic note collection and reporting. These reliability

procedures will include on-site training, parallel field observations, training sessions at Juárez and Associates in Los Angeles, and supervisorial monitoring of field data collection. The supervisor will systematically visit the participant researchers in the field for parallel observations during each field work session. The supervisor will thus be the on-site calibrator of observational data, collecting rating sheets, check lists, and other systematic assessment instruments. Instruments will be sufficiently simplified to permit regular and detailed calibration. Juárez and Associates will produce a report on the reliability of the ethnographic data and field observations at the conclusion of the project. (See Juárez and Associates, December 1978a, for a discussion of data quality monitoring and reactivity issues.)

B. Data Collection Activities

1. Assessment of Implementation

The evaluation of program implementation concerns the congruence between the proposed programs as written by the curriculum model developers and the programs as actually implemented at the replication sites: in other words, the relationship between words and deeds, or between that which is intended and that which is manifested. This objective will necessitate observations of actual behavior in the classroom: what teachers do, classroom arrangements, allocation of people, parental participation, materials

available, and so on. Without this information it will be difficult or impossible to inform others as to the nature, content, and staffing of the programs except insofar as they were originally proposed. Additionally, this information is also important for the interpretation of impact data. Without this information, also, impact data may be erroneously interpreted as if a program had been implemented as proposed when in fact it was not.

A second and related concern of the evaluation of program implementation has to do with the degree to which the various programs have been implemented. Here the issue is not whether the implemented program has any relationship to the proposed program, but instead the extent to which the program was implemented and the determinants of such implementation. Many constraints and obstacles to full implementation of any given model are likely to exist causing certain compromises to be made in the operation of the program. For example, the site may be inappropriately or poorly equipped, key personnel may be absent, on leave, or transient; the community may not desire the program; instructional materials may not arrive on time; and so on. Knowledge of these constraints and their influence on implementation of the program as intended is especially crucial to the dissemination of information to teachers, parents, and others who may subsequently seek to implement one of the models in their own settings.

In order to collect the information needed to assess implementation the following tasks must be completed:

(1) Develop Instruments.

(a) Develop a Model-Specific Implementation Checklist. A checklist of important classroom activities and materials necessary for their successful implementation will be developed. This checklist will be organized according to five categories of information which have been generated from a review of the implementation objectives of each of the curriculum models. The five categories of information include (1) the setting and material resources available, (2) the schedule and organization of the programs, (3) the individuals involved, (4) the attitudes of the individuals involved, and (5) the activities in which the individuals are engaged. (For a discussion of these five categories and how they were developed, see Juárez and Associates, December 1978a.)

(b) Review and Revision of the Checklist. Model-specific checklists will be submitted to the respective curriculum model developers for review and suggestions. Following this review, the checklists will be revised by the evaluators, taking into account the model developers' comments.

(c) Develop Teacher Questionnaires. Teacher questionnaires have been developed and submitted to the ACYF project officer for review. The questionnaires include sets of questions regarding the attitudes, knowledge, and behaviors of those teachers and teacher aides employed at the Head Start Centers.

(2) Use of Instruments to Collect Implementation Data. The assessment of implementation consists of three primary and interrelated data collection activities. They are as follows:

(a) Teacher Interviews. Interviews with teachers and teacher aides are scheduled to take place as early and as late in the program as possible. Information will be collected by means of self-administered, structured questionnaires which, as mentioned above, include questions regarding attitudes.

(b) Use of the Checklist. The implementation checklist will be used by the field workers to collect data four times during the evaluation year: at the beginning of the program, during December 1979, in February of the following year, and during the post-test in May or June 1980. These four data points were chosen because implementation is an ongoing process requiring a strategy that collects relevant information throughout the entire experimental period. However, in order to answer why each program is being implemented as it is at each of the four points in time, we will supplement the information gathered by means of the checklists with focused ethnographic notes.

(c) Focused Ethnographic Notes. In addition to the use of the model-specific implementation checklist four times during the academic year, ethnographic

observations which provide contextual descriptions of classroom activities will be conducted. These observations will be organized into a series of notes regarding some of the determinants of implementation to include general operation of the school, attendance, parental participation, and any other factors discovered which might adversely influence or facilitate implementation.

The combination of interviews, checklists, and notes will provide a comprehensive and on-going picture of the process of implementation at each of the four PR-intensive replication sites.

2. Assessment of Program Impact

An important goal of this evaluation is to accurately assess each model's impact on children enrolled in the program, on their parents, and on the Head Start teaching staff. The work statement of the evaluation calls for a randomized selection of students and families to treatment and control groups, and for pre- and post-test change measures to be administered. Some of the instruments selected by the evaluators to assess program impact will address the performance of Head Start and control group children on selected measures of social competence (Juárez and Associates, July 1978). However, a basic characteristic of our approach is to avoid exclusive reliance on pre- and post-testing as a measure of very young children's competencies (or program effectiveness) and instead to use several data gathering methods (see Juárez and Associates, December 1978a).

Therefore, in addition to the use of pre- and post-tests as measures of impact, we have developed additional strategies that include observations over time of children's behaviors in naturally occurring classroom contexts.

a. Assessment of Child Impact

The participant researcher will utilize a variety of data collection techniques in the field. Juárez and Associates has already reviewed and recommended psychometric instruments (Juárez and Associates, July 1978). In developing this document, observational instruments used in previous evaluation studies (e.g., the National Day Care Cost-Effects Study) were reviewed and several instruments and data gathering approaches were discussed with the consultant group. Our interest is in using overlapping methods which then can be applied to the assessment of program impact and implementation (see Appendix 3 for a conceptual model). In addition, we wanted to use procedures that were flexible in their application and that could take into account the shifting social contexts in which classroom behaviors occur as powerful determinants of performance. (For a discussion of this, see Juárez and Associates, December 1978a.) Thus, we propose that the duties of the participant researchers while in the school setting should be a mixture of (1) direct and systematic classroom observations, (2) weekly ratings of program-related dimensions (e.g., language use), and (3) qualitative notes focused on evaluation domains and program implementations.

The following tasks must be completed to assess the impact on the child.

(1) Select Test Battery. Tests were selected following procedures carefully developed in order to choose instruments of high technical quality which have a high degree of congruence with the curricular objectives to be evaluated (see Juárez and Associates, July 1978).

(2) Pilot Test Battery. Juárez and Associates is currently conducting a pilot study of parent and child impact measures and the administration and monitoring procedures that are to be used during the evaluation. The pilot study will take place at four of the eight sites where the curriculum models are being implemented. (For further information, see Juárez and Associates, December 1978b.)

(3) Time and Event Samples. Time and event samples of children in the classroom will be employed to monitor behaviors in various contexts. A record sheet will be used to code samples of children's behavior *in situ*. The primary purpose of this coding is to collect behavioral samples (e.g., descriptive language use) which are context-specific and relevant to the program objectives. These repeated observations of target behaviors in specific situations will provide a productive measure of what is happening within the classroom as the year progresses. (For a more extensive discussion and examples of this technique, see Juárez and Associates, December 1978a.)

(4) Observer Ratings. Direct classroom observations are a powerful but relatively difficult and complex method of data collection. Direct observations will be supplemented with a weekly rating of program-related domains (e.g., language use) for each site. Each target child and his or her classroom teachers will be rated at the end of each week of observations within the classroom. These ratings will provide important evidence of change and of the effectiveness of each model in producing change over time. (For examples, see Juárez and Associates, December 1978a.)

(5) Focused Ethnographic Notes. In addition to the focused event samples, and the more global ratings of teachers and children, the PRs will be keeping a series of notes on topics relevant to the overall functioning of the school and the implementation of each model. A series of topical domains or an index and mnemonic system will be developed for this purpose. These notes will be frequently and regularly checked by the supervisors and constant monitoring of note completeness and quality will be essential.

There will be general index note categories extending across all of the models and field sites, and other index categories specific to the programmatic aims of each model. Categories currently being considered for implementation across all sites include: (a) naturalistic usage of communication in the second language within the

classroom, (b) functional facilitation of this usage, (c) teaching styles within the class environment, (d) organization of the room on a daily basis, (e) activities being performed, (f) daily routine of pupils and teachers, (g) sociometric data -- including what individuals children associate and talk with on a regular basis and in what language, and (h) teacher language use. These will be indexed and monitored based on the context in which they are used.

These focused field notes, specific to domains of interest to the overall evaluation and to each model, should provide an important qualitative assessment of the functioning of each school, its implementation of the model, and any exogenous factors which are likely to have influenced either the in-class observations and ratings, the pre- and post-test measures, or the experimental arrangement.

(6) Pilot Test of the Observational Methods. The tryout of the three observation techniques will be divided into two phases:

(a) Initial Tryout by Juárez and Associates Staff.

Starting in March, the observational methods to be used in the evaluation will be piloted at some of the local Head Start Centers. This initial debugging will focus on the practicality of the methods, the techniques for establishing and maintaining reliability, the amount of training, and the

practice needed to effectively collect data and to establish a routine use of the instruments and methods.

(b) Tryout by Participant Researchers. Once the PRs have been selected, an important part of their training will consist of on-site experiences and involvement with local Head Start centers. Therefore, the same people who will be using the instruments in the field will be conducting the second pilot tryout. Emphasis will be on establishing reliability, the accurate reporting of results, and establishing rapport. These training and practice sessions will last for two months. (See Juárez and Associates, December 1978a, for details of the training plan.)

b. Assessment of Parental Impact

The evaluation contract specifies that data collected from the parents of children participating in the evaluation will be used to assess the impact of the program on parental attitudes and knowledge regarding:

- Education in general
- Bilingual education in particular
- Educational and vocational aspirations for their children

In addition, daily-routine information will be collected from treatment and control group parents. The fundamental reason for collecting information on treatment and control group parents and families is that many of the families not selected for the

Head Start Center are expected to seek alternative preschool day-care or other education experiences for their children. Families may also expose their children to educational television, to several language environments, to caretakers, and so forth. We feel very strongly that it is important to monitor such data during the year. When the control group results are compared to the Head Start group after the post-tests are completed, Juárez and Associates will then be in a far better position to account for, interpret, and understand whatever similarities or differences may emerge between the two populations. Thus, these data are being collected to aid in understanding and interpreting results.

The two primary data collection methods to be used follow, along with a description of each.

(1) Parent Questionnaire. The parent questionnaire is designed to collect data on the parents of the children enrolled in the eight Head Start bilingual bicultural curriculum model sites and the parents of the control children in the evaluation. It is anticipated that a sample of forty-five parents from each control/experimental group will be administered the questionnaire in a pre- and post-test fashion. The questionnaire has been reviewed by the curriculum model developers and is currently being pilot tested on a representative sample of the target population for which it is intended.

(2) Phone-Call Updates. The treatment and control groups will be followed through a telephone updating procedure during the course of the evaluation year, between the

parental pre- and post-test interviews. A randomly selected subsample of fifteen families from the four PR-intensive sites will be called approximately every three months and information will be gathered on a number of topics directly related to home language use; exposure to other educational and preschool experience; change, mobility, and variability in residence and employment; and child health. The calls to families at "implementation-only" sites will be made quarterly. This is an inexpensive method of obtaining basic and important information on the out-of-school experiences.

The tasks necessary to complete the assessment of the impact on the parents are listed below.

(1) Prepare Checklist. A checklist of information needed to monitor out-of-school activities during the year will be prepared in April and piloted with a sample of Head Start parents in Los Angeles.

(2) Review Checklist. This checklist will be submitted for review to the consultants and the curriculum model developers, and revised according to the formative feedback received.

(3) Select Stratified Subsample. A stratified subsample of 15 control group families to be followed during the year at each evaluation site will be selected in August 1979 and the families will be contacted to obtain written consent for their participation in the monitoring activities.

c. Assessment of Teacher Impact

The evaluation contract specifies that the Head Start teachers' and aides' performances will be evaluated on measures designed to assess their understanding of the terms "bilingual" and "bicultural" in the context of an early childhood program, their attitudes toward children, their sensitivity to special ethnic and linguistic characteristics of children, and their ability to use these characteristics in a positive fashion in the instructional program. In addition, the teachers' degree of satisfaction with the curriculum models must be assessed. As mentioned earlier, the teachers' activities in the classroom will also be taken into account in the assessment of program implementation and of program impact.

The following tasks relate to the assessment of teacher impact.

(1) Development of Teacher Questionnaire. This instrument has been developed and is currently being reviewed by ACYF.

(2) Administration of Questionnaire. The questionnaire will be administered to all of the teachers and aides participating in the program as early and as late in the evaluation year as possible. Arrangements are currently being discussed.

(3) Observer Ratings. The direct classroom observations discussed in the "Child Impact" section will be supplemented with weekly ratings of program-related domains for each of the PR-intensive sites. Classroom teachers

will be rated at the end of each week of observations within the classroom. Rating categories include the use of both languages in the classroom, general classroom organization, and how parents are involved in the daily classroom activities.

(4) Ethnographic Notes. In addition to the ratings, the field observers will be keeping a series of notes on topics relevant to the overall functioning of the program. Teacher categories will include how he or she facilitates the use of the second language for communicative purposes, teaching styles and how they are related to those teacher behaviors specified in the curriculum, and any other exogenous variables that may have an impact on the teacher's performance.

V. Data Management

The importance of careful, continuous, and rigorous data management to the overall success of the evaluation should again be emphasized. As included in the Time Line of Activities (see Appendix A), the field supervisor will maintain constant contact with the field workers, make frequent site visits to conduct retraining and to check on reliability, and organize visits from the fieldworkers to Los Angeles for debriefing and recalibration.

It should be clear that in addition to the constant monitoring of activities, Juárez and Associates will process field notes, observations, and implementation reports as they are received on a regular basis by the field supervisor. These notes will be indexed and prepared for preliminary analysis as soon as feasible after receipt. This would enable us to assess whether the information collected is what is needed to satisfy the evaluation objectives. Necessary changes in the tasks of the field workers or in the focus of their tasks will be made as a result of these preliminary analyses and discussions. (For further information on data management, see Juárez and Associates, December 1978a.)

VI. Data Analysis

The data analysis for the evaluation combines both (1) psychometric data and (2) certain data from the observations in the form of ratings, time estimates, and frequencies of events. The analysis will proceed at two levels. At the first level, ethnographic observations will be combined with psychometric test data as a multi-index of the dependent variables of the study, such as second language comprehension and production. Both types of data will also be used to acquire a better estimate of the effects of the treatments from pre- to post-tests. The second level of analysis involves the use of data from the implementation component to interpret the results of the evaluation. For example, data from the implementation checklist can be used to determine what components of a given model have been implemented and the possible effects of non-implementation or partial implementation on the dependent variables. For further details about the analysis of the data, see Juárez and Associates, "Evaluation and Analysis Plan," 1978, and Juárez and Associates, December 1978.

APPENDIX A

TIME LINE OF ACTIVITIES FOR THE EVALUATION

1979	Activity
February 1-28	- Begin pilot testing of psychometric instruments at four curriculum sites.
	- Begin recruitment of Participant Researchers and Supervisor.
	- Begin recruitment of Data Manager.
March 1-31	- Hire support staff.
	- Hire Data Manager.
	- Develop checklist for phone-call updates.
	- Begin development of implementation checklist.
	- Begin field testing of observational instruments.
	- Begin development of plan for quality control of data from classroom observations.
	- Complete pilot testing at four curriculum sites.
April 1-30	- Complete implementation checklist and send to curriculum developers for review.
	- Pilot test phone-call update procedures on local families.
	- Arrange preschool sites for PRs to pilot-test instruments.
	- Continue field testing of observational instruments.
May 1-31	- Hire Supervisor of PRs.
	- Complete plan for quality control.
	- Revise implementation checklist.
	- Submit phone-call update procedures to curriculum developers for comments.
	- Continue pilot testing of procedure and instruments.

June 1-30

- Hire four PRs.
- Revise phone-call update procedures based on feedback from curriculum developers.
- Begin training of PRs.
- Develop data filing system for coding observations.
- Pilot-testing of all procedures and instruments by PRs.

July 1-31

- Continue training of PRs and Supervisor.
- Review results from pilot-testing.
- Submit report on selection of observation instruments to project officer (7/31/79).

August 1-31

- Place PRs on site.
- Conduct on-site orientation for PRs.
- PRs start ethnographic notes and observations.
- Select stratified sample of families from control and Head Start children.
- First progress report due from PRs to Juárez and Associates.
  - includes ethnographic notes, rating, and observations.

September 1-30

- Feedback to PRs from Supervisor regarding observations and notes.
- First site visit by Supervisor for recalibration and reliability check.
- Pre-testing of children with psychometric instruments.
- First data point for implementation assessment at non-intensive sites.
  a. Conduct initial teacher interviews.
  b. Fill out checklist on implementation.

- Progress report due from PRs to Juárez and Associates.

- Materials from PRs are categorized, indexed, and filed. (ongoing)

October 1-31

- Feedback to PRs from supervisor regarding observations and notes.

- Collect time and event sample of classroom activities.

- Categorize, index, and file materials from PRs.

- Progress report due from PRs to Juárez and Associates.

November 1-30

- Feedback from Supervisor to PRs.

- Second visit by Supervisor for recalibration and reliability check.

- Second data point for implementation assessment at non-intensive sites.

  a. Fill out checklist on implementation.
  b. Progress report due from PRs to J & A.

- Categorize, index, and file materials from PRs.

December 1-31

- Feedback from Supervisor regarding observations and notes.

- Collect time and event samples of classroom activities.

- Progress report due from PRs to Juárez and Associates.

- Categorize, index, and file materials from PRs.

1980	Activity
January 1-31	- Conduct debriefing sessions for PRs in Los Angeles. a. Recalibrate PRs. b. Conduct paired observations for reliability check for PRs.
February 1-29	- PRs at site. - Collect time and event samples of classroom activities. - Progress report due from PRs to Juárez and Associates.
March 1-31	- Feedback to PRs from Supervisor. - Third data point for implementation assessment. a. Conduct teacher interviews. b. Fill out checklist on implementation. - Third site visit by Supervisor. - Progress report due from PRs to Juárez and Associates. - Categorize, index, and file materials from PRs.
April 1-30	- Feedback to PRs from Supervisor. - Progress report due from PRs to Juárez and Associates. - Categorize, index, and file materials from PRs.
May 1-31	- Feedback to PRs from Supervisor. - Begin post-test of psychometric instruments. - Collect time and event sample of classroom activities.

- Fourth site visit by Supervisor.

  a. Recalibration.
  b. Reliability check.

- Fourth data point for the implementation component:

  a. Conduct teacher interviews.
  b. Fill out implementation checklist.

- Categorize, index, and file materials from PRs.

June 1-30

- Post-testing completed.

- Second debriefing session for PRs begins in Los Angeles

July 1-31

- Debriefing session of PRs continues.

- Preliminary report to ACYF on debriefing session.

August 1-30

- Draft Preliminary Results report.

SUMMARY OF DA... COLLECTION ACTIVITIES

Assessment of Child Impact[1]

Domain	Instruments/ Data Source	Respondent(s)	Sample Size	Frequency	Unit of Analysis
Spanish/English Language Production	1. Tests	H.S. & Control	45 H.S., 45 Cont.	Pre & post	Child,class,site
	2. Event samples	H.S.	H.S. Subsample 15	Quarterly	Child
	3. Observer ratings	H.S.	H.S. Subsample 15	Quarterly	Child
	4. Ethnographic notes	H.S.	H.S. Subsample 15	Quarterly	Child
Spanish/English Language Comprehension	1. Tests	H.S. & Control	45 H.S., 45 Cont.	Pre & post	Child,class,site
	2. Event samples	H.S.	H.S. Subsample 15	Quarterly	Child
	3. Observer ratings	H.S.	H.S. Subsample 15	Quarterly	Child
	4. Ethnographic notes	H.S.	H.S. Subsample 15	Quarterly	Child
Concept Development	1. Tests	H.S. & Control	45 H.S., 45 Cont.	Pre & Post	Child,class,site
	2. Event samples	H.S.	H.S. Subsample 15	Quarterly	Child
	3. Observer ratings	H.S.	H.S. Subsample 15	Quarterly	Child
Socio-emotional Dev.	1. Observer ratings	H.S.	H.S. Subsample 15	Quarterly	Child
	2. Ethnographic notes	H.S.	H.S. Subsample 15	Quarterly	Child
	3. Teacher ratings	H.S.	45 H.S.	Quarterly[2]	Child
	4. Phone call updates	Control	Subsample 15	Quarterly	Site

[1] This table represents the evaluation activities at the P.R.-intensive replication sites. The quarterly ratings and reports consist of two weeks of daily classroom observations.

[2] Varies with how often individual models require teachers to rate the students.

Assessment of Parent Impact

Domain	Instruments/ Data Source	Respondents	Sample Size	Frequency	Unit of Analysis
I. Background Characteristics	Parent Questionnaire	Parents of H.S. & Control Children	45 Control 45 Head Start	Pre & Post (some items)	Site, child
II. Parental Attitudes Toward:					
a. Education in general	Parent Questionnaire & Phone-call updates	Parents of H.S. & Control Children	45 Control 45 Head Start	Pre & Post	Family[1]
			Subsample: 15 Control 15 Head Start	Quarterly	Family
b. Bilingual Education	Parent Questionnaire & Phone-call updates	Parents of H.S. & Control Children	45 Control 45 Head Start	Pre & Post	
			Subsample: 15 Control 15 Head Start	Quarterly	Family
c. Biculturalism	Parent Questionnaire	Parents of H.S. & Control Children	45 H.S. Control 45 Head Start	Pre & Post	Family
d. Educational & Vocational aspirations for their children	Parent Questionnaire	Parents of H.S. & Control Children	45 Control 45 Head Start	Pre & Post	Family
e. Attitudes toward Head Start Project	Parent Questionnaire & Phone-call updates	Parents of H.S. children only	45 Control 45 Head Start	Pre & Post	Family
			Subsample: 15 Head Start	Quarterly	Family

[1]Family level: Data sources for these variables will include the parents and other members of the child's family. Our intention for collecting information on these variables is to assess the effect of the treatment on the child's home environment.

omain	Instruments/ Data Source	Respondents	Sample Size	Frequency	Unit of Analysis
II. Involvement with Head Start Program	Observer ratings Ethnographic notes	Parents of H.S. children only	Subsample: 15 Head Start children's parents	Quarterly	Classroom, site
V. Activities in Home					
a. Use of Language by Children & Adults	Parent Questionnaire & Phone-call updates	Parents of H.S. & Control Children	45 Control 45 Head Start Subsample: 15 pairs control 15 pairs H.S.	Pre & Post	Family
b. Reading Activities	Phone-call updates	Parents of H.S. & Control Children	Subsample: 15 pairs control 15 pairs H.S.	Quarterly	Family
c. Formal & Informal Educational Activities of Parents	Phone-call updates	Parents of H.S. & Control Children	Subsample: 15 control 15 Head Start	Quarterly	Family
d. Formal & Informal Educational Activities of Children	Phone-call updates	Parents of H.S. & Control Children	Subsample: 15 control 15 Head Start	Quarterly	Family

Domain	Instruments/ Data Source	Respondents	Sample Size	Frequency	Unit of Analysis
e. Amount of Time Child Watches T.V. & Which Language	Phone-call updates	Parents of H.S. & Control Children	Subsample: 15 control 15 Head Start	Quarterly	Child
f. Use of Social Service Agencies in the Community	Phone-call updates	Parents of H.S. & Control Children	Subsample: 15 control 15 Head Start	Quarterly	Family

Assessment of Teacher Impact

Domain	Instruments/ Data Source	Respondents	Sample Size	Frequency	Unit of Analysis
I. Background Characteristics	Teacher Questionnaire	Teacher & aides	Total Population at each site	Twice a year	Classroom, site
II. Understanding of bilingualism & biculturalism	Teacher Questionnaire	Teacher & aides	Total Population at each site	Twice a year	Classroom, site
III. Attitudes of teachers towards:					
a. Spanish dominant children	Teacher Questionnaire	Teacher & aides	Total Population at each site	Twice a year	Classroom, site
b. Bilingual H.S.children	Teacher Questionnaire	Teacher & aides	Total Population at each site	Twice a year	Classroom, site
c. Parents of H.S. children	Teacher Questionnaire	Teacher & aides	Total Population at each site	Twice a year	Classroom, site
d. Curriculum model	Teacher Questionnaire	Teacher & aides	Total Population at each site	Twice a year	Classroom, site
e. Sensitivity to children's special characteristics	Teacher Questionnaire	Teacher & aides	Total Population at each site	Twice a year	Classroom, site

Domain	Instruments/ Data Source	Respondents	Sample Size	Frequency	Unit of Analysis
IV. Teacher classroom activities:					
a. Ability to incorporate children's special characteristics in the teaching process	1. Teacher Questionnaire 2. Ethnographic notes	Teacher & aides	Total Population at each site	Twice a year Quarterly[1]	Classroom, site
b. Teaching styles within the class environment (model-specific requirements)	1. Observer ratings 2. Ethnographic notes	Teacher & aides	Total Population at each site	Quarterly Quarterly	Classroom, site
c. Organization of the room & activities being performed	1. Observer ratings 2. Ethnographic notes	Teacher & aides	Total Population at each site	Quarterly Quarterly	Classroom, site
d. Daily rountine of pupils & teachers	1. Observer ratings 2. Ethnographic notes	Teacher & aides	Total Population at each site	Quarterly Quarterly	Classroom, site
e. Language use in the classroom	1. Observer ratings 2. Ethnographic notes	Teacher & aides	Total Population at each site	Quarterly Quarterly	Classroom, site

[1]At the four P.R.-intensive replication sites, the quarterly ratings and reports consist of two weeks of daily classroom observations.

Bibliography

Juárez and Associates, Inc. Review and recommendation for the test battery. An evaluation of the Head Start bilingual bicultural curriculum development project. Report prepared for the Administration for Children, Youth and Families, Contract No. HEW-105-77-1048, July 1978.

Juárez and Associates, Inc. A qualitative/quantitative data gathering approach. An evaluation of the Head Start bilingual bicultural curriculum development project. Report prepared for the Administration for Children, Youth and Families, Contract No. HEW-105-77-1048, December 1978 a.

Juárez and Associates, Inc. Phase II. A plan for the pilot study of child and parent impact measures. An evaluation of the Head Start bilingual bicultural curriculum development project. Report prepared for the Administration for Children, Youth and Families, Contract No. HEW-105-77-1048, December, 1978 b.